~~[illegible]~~
804

ENVIRONMENTAL RESEARCH

RESEARCH AND TECHNOLOGY BRANCH

AN EXPERT SYSTEM FOR WATER QUALITY ASSESSMENT OF ONTARIO RIVERS

R. A. C. PROJECT NO. 332 PL
WatQUAS 2.0

Environment Ontario
Jim Bradley, Minister

ISBN 0-7729-6048-8

AN EXPERT SYSTEM FOR WATER QUALITY ASSESSMENT OF ONTARIO RIVERS

R. A. C. PROJECT NO. 332 PL
WatQUAS 2.0

Report prepared for Environment Ontario by:

T. E. Unny (Principal Investigator
and
W. C. Allison

Systems Design Engineering
University of Waterloo
Waterloo, Ontario N2L 3G1

SEPTEMBER 1989

RECYCLABLE

DISCLAIMER

Abstract

WatQUAS 2.0: An Expert System for Water Quality Assessment of Ontario Rivers

Intelligent Knowledge Based Systems are computer programs which exhibit machine intelligence. Machine intelligence is the capability of a computer to efficiently search through large quantities of heuristics (rules of thumb), and expert and domain knowledge in order to achieve inferential conclusions. WatQUAS 2.0 is an Intelligent Knowledge Based System (Expert System) for the assessment of water quality of Ontario rivers. WatQUAS 2.0 operates on an IBM PC compatible computer and is highly user interactive. A Data Base Management System is utilized to organize and contain large quantities of historical water quality data, parameter and site specific knowledge. WatQUAS 2.0 contains knowledge pertaining to approximately 255 water quality contaminants. The Expert System component of WatQUAS 2.0 examines various water quality problems and situations and achieves inferential interpretations and conclusions. The water quality assessment techniques employed by WatQUAS 2.0 have been expanded and enhanced from the prototype version. Future work involves completing the computer programming of the Expert System, expanding the knowledge base and programming WatQUAS to examine more water quality assessment areas. Comprehensive testing and evaluation is also required.

Acknowledgments

I would like to thank Professor T.E. Unny for his support and commitment to the project. I would also like to thank Professor G. Farquhar for his support, specifically for the acquisition of the Expert System Language compiler.

I would also like to thank Drs. Lloyd Logan and Byron Bodo and Mr. Brian Whitehead for their advice in developing WatQUAS 2.0.

Steadman Chui is also to be thanked for his help in programming in the C language.

TABLE of CONTENTS

LIST of TABLES

LIST of ILLUSTRATIONS

1.0 Introduction

The term EXPERT SYSTEM has been applied indiscriminately to many diverse types of computer programs. It has become a "catch-all" phrase for any software that provides the user with more than a numerical response to a problem. The label, EXPERT SYSTEM, has evolved into a cliche that is over-used and perhaps not well understood.

"Machine Intelligence" distinguishes true Expert Systems from deterministic computer programs. "Machine Intelligence" is the ability of a computer to reach complex inferential solutions to problems. The computer must be capable of searching through many heuristics (rules of thumb) and selecting the appropriate rules that suits each individual situation. A more suitable title for an Expert System which exhibits machine intelligence is an Intelligent Knowledge Based System (IKBS). This name clearly implies that the system contains knowledge that emulates the information stored in the human brain. Searching through a large array of heuristics is the computer equivalent to the human thought process.

WatQUAS 2.0 is an Intelligent Knowledge Based System for the assessment of water quality in Ontario rivers. A comprehensive numerical analysis is conducted on the historical water quality record of a river monitoring site.

An expert interpretation of the water quality at the site is completed by utilizing the results from the numerical analysis, a large knowledge base and an inferential engine. Conclusions regarding the origins, seriousness and possible solutions to the water pollution problems are presented by WatQUAS 2.0. The water quality assessment components, Expert System operation modules and expansion of the knowledge base has been completed. There remains work to be completed linking together the modules and optimizing the software package. A graphics package must be installed in WatQUAS 2.0 and the "recognize-act" cycle of the expert system must be converted to IBM PC format.

1.1 Scope of Thesis

This thesis deals with the development of WatQUAS 2.0. The expansion of the knowledge base and enhancement of the water quality assessment techniques is the main focus of this work. Chapter 2 presents a brief study of knowledge engineering and knowledge extraction techniques utilized by Expert System developers. A review of water quality indices and flow weighted pollutant load estimators are also contained in this chapter. These two areas are important for quantifying and assessing water pollution problems.

The third chapter of this thesis discusses the prototype

Expert System, WatQUAS 1.0. An overview of WatQUAS 1.0 is presented which outlines how it works. A critique of the prototype Expert System identifies the problems and shortfalls inherent in WatQUAS 1.0. Chapter 4 describes the development and components of WatQUAS 2.0 for the micro-computer. The water quality assessment techniques utilized by the second version of the Expert System are thoroughly discussed. The methodology for conducting an expert interpretation of the water quality analyses by WatQUAS 2.0 is also presented.

The fifth chapter of this thesis describes how knowledge extraction techniques have been applied specifically to WatQUAS 2.0. The expansion of the knowledge base and the heuristics of the Expert System are also outlined. Chapter 6 discusses future work and recommendations for the development of the WatQUAS Expert System.

Finally, conclusions regarding WatQUAS 2.0 and a summary of of the completed work to date are presented in Chapter 7. WatQUAS 2.0 is a relatively small and incomplete Expert System, continual work and development is still required to make it a comprehensive and versatile water quality assessment tool.

2.0 Knowledge Engineering

Knowledge engineering is a high technology field that emerged from new developments and advances in Artificial Intelligence. A knowledge engineer is responsible for the task of capturing knowledge and storing it in a form that is readily usable by an Intelligent Knowledge Based System (IKBS). In this section, a brief background and definition of knowledge engineering will be presented. Knowledge engineering techniques and knowledge extraction methods for water quality assessment will also be discussed. Finally, methods for storing and organizing the expert knowledge will be examined.

2.1 Knowledge Engineering Background

The most important component of an IKBS is the knowledge block. Construction of a knowledge base is referred to as knowledge engineering. This is the process of incorporating relevant information and data into an organized knowledge base for use by an Expert System. There are mainly two types of knowledge that are required by an Expert System. There is domain knowledge, this consists of accepted knowledge which can be obtained from books, journals, manuals, etc.. This is the easiest type of knowledge to deal with because it is not controversial and is usually not redundant. The knowledge engineer screens large

quantities of domain knowledge and incorporates the relevant information into the IKBS.

The second type of knowledge required by an Expert System is heuristic knowledge. This type of knowledge is composed of "rules of thumb" that are elicited from experts in the field. Usually experts have gained their knowledge through training and many years of practical experience. The knowledge engineer is responsible for extracting knowledge from the expert and translating it into an Expert System usable format. The interaction between the knowledge engineer and domain expert is crucial. The knowledge engineer must guide the expert through many hypothetical situations and "what if" cases in order to extract relevant information.

The heuristics for a given situation, that are elicited from experts may vary from one expert to another. The knowledge engineer can solve this problem by programming the IKBS to respond with multiple answers or by assigning confidence weights to each expert's opinion. The heuristic with the largest weight is the one given highest priority by the Expert System. Conflicting heuristics are desirable in an Expert System because they illustrate to the user that there is controversy surrounding a subject. They also diminish the often mistaken concept that the Expert System

is infallible and always produces the ultimate answer for all situations. By producing multiple solutions the Expert System shows the user that even recognized experts can not reach a consensus on a solution to the problem. Hopefully, this technique will discourage users from blindly adhering to the solutions produced by the IKBS.

2.1.1 Logic Programming

Present day computers are not very efficient at understanding everyday human language. The knowledge engineer must translate the expert information (domain and heuristics) into a format understandable by an Expert System. Logic programming is often used by knowledge engineers to communicate effectively with a computer. Logic programming is based on predicate calculus, complex symbols are used to form a language which is exact and not redundant [Tore 1987]. Programming in logic is the most efficient way for humans to interface with computers. For a thorough discussion of predicate calculus see [Allen 1986].

2.1.2 Fuzzy Logic and Bayesian Probability

One of the major problems in Expert System development is programming the system to examine the uncertainty of a solution or answer. Human experts often state their

responses in uncertain terms, outcomes of events are labeled as "likely", or "probable". It is extremely difficult to program a computer to respond in uncertain or "fuzzy" terms. Individuals have unique interpretations of the degree of uncertainty associated with "fuzzy" terms. For example, it is impossible to assign a probability to the term "likely". The knowledge engineer does not know if "likely" means 60%, 70% or 80% probability of the event occurring.

Certainty factors have often been used to indicate the confidence in an interpretation. The factors usually range from (0 - 1.0), with 1.0 indicating complete confidence in a solution. Fuzzy set theory has been applied to many problems where rigid logic and quantitative mathematics are inappropriate because of the inherent uncertainty. The logic programming described in section 2.2.1 contained propositions which were either true or false. In fuzzy set logic a proposition has a degree of truth associated with it [Hart 1986 p. 102]. A fuzzy set is described by;

$$F = M_1/U_1 + M_2/U_2 + \ldots + M_n/U_n$$

Where; F = Fuzzy Set,

U_i = the set of all possible outcomes,

M_i = the degree of membership in the fuzzy set of each outcome,

$+$ = denotes union.

For water quality assessment the magnitude of the degree of membership for each outcome is assigned by an expert based upon personnel experience. The expert realizes that certain outcomes are possible given a particular pollution situation. A numerical value based upon the likelihood of the event occurring in relation to other events is assigned by the expert.

Fuzzy set logic has been widely applied in many expert systems. However, in water quality management it is often difficult to assess all the possible outcomes from a situation where water pollution has occurred. It is even more difficult to assign a degree of membership (likelihood) to the various events. The unpredictable nature of water quality and the usual sparse information available in many situations makes fuzzy set theory difficult to apply to water quality.

Bayesian theory is a technique often used to predict the posterior probability of uncertain events. Bayes theory [Bunn 1984] predicts this probability given the prior probability of the event occurring.

$$P(S|C) \propto P(C|S)P(S)$$

where; $P(S|C)$ = the posterior probability of S given C,

$P(C|S)$ = the probability of C conditional upon the assumption S occurring,

P(S) = the prior probability of S.

The prior belief is transformed into a posterior probability according to the strength of the prior belief and the likelihood of the data utilized to form this belief that the event will occur [Bunn 1984 p. 117].

An example related to water quality management would be the calculation of the probability of future water quality violations occurring in a stream. The prior probability is based upon the historical quality record of the stream, that is the number of previous violations that have occurred. The likelihood function is derived from the confidence that exists in the historical quality record. Combining these two terms yields the posterior probability of additional violations occurring.

For water quality assessment, application of Bayesian theory is difficult and can often produce misleading results. The main difficulty lies in attempting to compute the prior probabilities of a given event. Samples of water quality data are usually discrete, 20 - 30 readings a year. Attempting to compute probabilities without a continuous data record can often lead to inaccurate estimates. For example, assessing the probability of stream quality violations using a record consisting of 30 samples for each of 3 years would likely yield an uncertain answer.

Although it may be possible to program an Expert System to include uncertainty, it is difficult to incorporate judgment into the software. Complex heuristics are often used to imitate human judgment. However, the knowledge engineer cannot foresee all possible situations where judgment may be required.

In water quality management an area where judgment forms a criteria in decision making is the rigidity of numerical values. For example, a Provincial Water Quality Objective (PWQO) for a pollutant may be 5.0 mg/l. A water quality sample analyzed for this pollutant may contain 5.1 mg/l. Then, is this concentration level acceptable? Some factors which may affect the decision are;

* Is the pollutant toxic?
* What are the water uses?
* What is the history of the pollutant at the site?
* Is there a threat to aquatic life?

Clearly, some form of judgment is required by the computer for this situation. The computer must be able to distinguish between situations when 5.1 mg/l is acceptable and when it is not. A simplistic method of handling the above example is to incorporate an allowance on both sides forming a range for the PWQO;

PWQO = 5.0 mg/l + - X%

The problem with this solution is that the "X%" becomes a rigid number. The identical problem is then encountered with concentration levels marginally outside the PWQO range. Incorporating "human-like" judgment into Expert Systems requires extensive research and development.

2.1.3 Knowledge Organization

Although a large and comprehensive knowledge base is an integral part of an Expert System, equally vital is that the knowledge base be organized. The Expert System and the users must have direct access to the knowledge base. This is important for locating, modifying and expanding the information in the knowledge base. Data Base Management Systems (DBMS) are ideal tools for organizing and managing knowledge blocks of Expert Systems.

2.2 Knowledge Engineering for Water Quality Assessment

This section contains a discussion of two methods of knowledge engineering for water quality assessment. Water quality indices are described and reviewed in the first part of this section. This will form the basis for the construction of a new index for application in WatQUAS 2.0. The theory of flow weighted pollutant load estimates using ratio estimators is contained in the second part of this

section. The use of ratio estimators is the technique recommended by the Ontario Ministry of the Environment (MOE) for pollutant load estimating.

2.2.1 Water Quality Indices

Water pollution problems at a specific site are often difficult to assess in terms of overall effects and seriousness. Water quality experts are frequently asked to simplify complex pollution problems into a form which may be intuitively understood by the non-expert. A water quality index is a form of this simplification which condenses information regarding complex water pollution problems into a single number. An index which can condense many different problems into a single number is an ideal tool for utilization by an Expert System.

2.2.1.1 Critical Review of Water Quality Indices

This section examines some generally accepted water quality indices in order that the best features from each may be utilized to produce a comprehensive and robust water quality index for WatQUAS 2.0.

Water quality indices can be divided into two broad categories; indices which consider individual pollutants and indices which examine the site as a whole without recognizing specific contaminants. Each method has

advantages and disadvantages associated with it and these make neither type entirely acceptable.

2.2.1.2 Pollutant Specific Indices

Water quality indices which examine individual parameters usually utilize some form of rating curve to "score" each individual pollutant. The rating curve links the concentration or quality measurement (for instance; temperature is measured in degrees) of the parameter with the quality of water at the site. These rating curves are usually expressed in the form of graphs or mathematical equations. Their purpose is to transform some measure of the parameters "in-stream" quality into a non-dimensional number. This transformation eliminates the units associated with each parameter, which often differ and is the major cause of difficulty with aggregating the pollutants.

The rating curves are usually based upon various experts' opinions concerning the effects and seriousness of the individual pollutant at different levels. The "delphi" technique of pooling experts opinions appears to be the most widely accepted and used method. [Dinius 1987], [Couillard 1985] and others have used this technique to construct the rating curves utilized by the water quality indices developed by them.

There is conflicting opinion in the technical literature as to what measure of the concentration or quality of a parameter should be utilized in the rating curve process in order to determine a score. The same measure must be utilized for the entire data set regardless of the nature of the data. Some experts contend that the arithmetic mean of the time series history of a parameter is the best indicator to use. Other experts prefer the geometric mean of the pollutant time series. Water quality data are frequently skewed to the right, this causes the arithmetic mean to be biased high and would cause the water quality index to be artificially lowered [Bodo 1988]. Utilizing the geometric mean circumvents this problem by introducing a log transformation of the data to eliminate the skew.

Regardless of whether arithmetic or geometric means are utilized, neither method considers frequency, duration or magnitude of violation. Extremely high pollution levels are compensated by low levels in the parameter aggregation formula when means are utilized.

A major problem with rating curves is that they are not responsive to the sensitivity of an individual site. For example, low levels of phosphorous pollution may be more hazardous to the environment at one site than high levels of phosphorous at another site. Site sensitivity depends

upon the ambient conditions at each location.

Very few of the common indices contain rating curves for more than a dozen parameters. The largest number of rating curves contained in any index was found to be 72 [Couillard 1985]. Most water quality indices utilize only a limited number of "critical" parameters to compute the quality. The "critical" pollutants are chosen by experts and are applied universally at every site, regardless of local conditions and specific problems. A stream may contain numerous hazardous contaminants, not specifically addressed by the index. If water quality indices only consider a few pollutants, the index could indicate a good water quality situation when in reality there are serious pollution problems.

After the individual scores are assessed for each parameter, they must be aggregated to form the overall water quality index for the site. Before combining, the individual parameters may be weighted. Weighting is a means of indicating the importance of a parameter in relation to other parameters. The larger the weighting factor assigned to a parameter (with weights summing to unity) the more critical the pollutant. The individual pollutants can either be aggregated by summation or multiplication. Table 2.1 shows the various aggregation

Table 2.1 Aggregate Methods

(Couillard 1985)

Method	Equation
Unweighted Sum	$I = \frac{1}{n} \Sigma_{i=1}^{n} q_i$
Weighted Sum	$I = \Sigma_{i=1}^{n} q_i w_i$
Unweighted Product	$I = (\Pi_{i=1}^{n} q_i)^{\frac{1}{n}}$
Weighted Product	$I = \Pi_{i=1}^{n} q_i^{w_i}$

where;

I = water quality index

q_i = individual parameter score

w_i = parameter importance weight

n = number of pollutants

methods generally used in water quality indices.

Ambiguity is a problem when the unweighted sum method is used. This occurs when the aggregated unweighted sum exceeds a critical limit value but none of the individual scores exceed the critical value.

Eclipsing is often a problem when the summation or the product aggregation techniques are used to combine scores. This is a situation when the overall index is satisfactory, however, one or more of the individual pollutants are a problem. One method which eliminates this problem of eclipsing is to utilize only the lowest scored parameter in the overall water quality index:

$$I = \min(q_1, q_2, \ldots q_n)$$

where; I = water quality index,

q_i = individual parameter score.

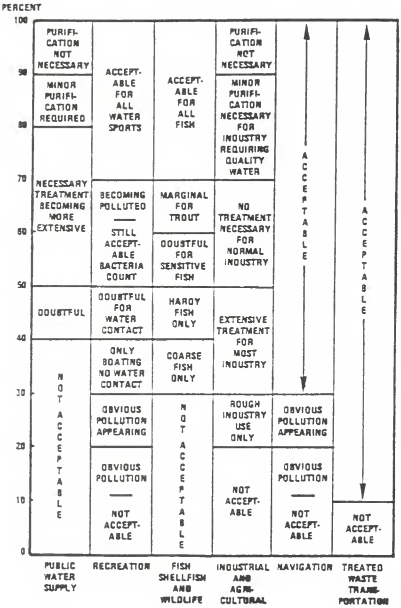
This technique does not indicate the overall water quality situation at the site because it considers only the worst quality pollutant.

The method generally recognized as the best for parameter aggregation is the weighted product technique. The weighting is recommended when there are a finite number of pollutants and their effects well known. The multiplicative aggregation is recommended because it

consistently produces a water quality index which is lower than or equal to that from the weighted summation technique. This method eliminates over-estimation of the water quality at a site.

The last step in calculating an index is to interpret the numerical value. A verbal rating system is used to describe the quality of water at a site. Figure 2.1 shows a rating system suggested by [Dinius 1987] which also accounts for water use at the site.

An area of concern for people wanting to utilize this type of water quality index in Canada is whether the parameter weights and rating curves are transferable to Canada from other countries of origin. Most of the established indices have been developed in the United States for American rivers. DO and BOD are typically weighted stronger for American rivers than would be required for Ontario rivers because of the more temperate climate encountered in the south. In colder climates DO and BOD lose the high priority rating assigned in most American water quality indices [Couillard 1985]. Any water quality index developed outside Ontario should be examined closely before being transferred to the province.



[Dinius 1987]

Figure 2.1 Water Quality Index Descriptions

2.2.1.3 Site Specific Indices

The second type of water quality index considers the water quality at the site as a whole, without examining the contribution of individual pollutants. The basic method was developed by the MITRE Corporation for the United States Environmental Protection Agency. It is essentially an index which indicates pollutant severity at a site. A common name of this index is the "PDI index" (prevalence, duration, and intensity of pollution over an area) [Truett 1975].

Prevalence (P) represents the length of the stream which does not meet water quality objectives, it is expressed in terms of a length. There is no distinction made as to which pollutants exceed the stream standards, only that the stream does not meet quality criteria.

The duration (D) of the pollution problem is rated in terms of the number of seasons in a year in which violations have been recorded. The following weights have been assigned to indicate the number of seasons in which there is a problem:

* .4 for any violations occurring in a single season,
* .6 for any violations occurring in two seasons,
* .8 for any violations occurring in three seasons,
* 1.0 for any violations occurring in all four seasons.

A violation of any pollutant in any season counts as a site violation, the index does not distinguish between which pollutants have been in violation.

The severity of the pollution is accounted for by utilizing an intensity factor (I). This index is expressed in terms of the effects of the pollutant on each specific site rather than parameter specific information. The effects are divided into three categories, the summation of the maximum weight from each category equals 1.0. The three categories and the breakdown of weights from each are listed in table 2.2.

The overall index is the product of the prevalence, duration and intensity divided by the total length of the stream (same units as P) [Truett 1975];

$$V = (P * D * I) / M$$

where; V = water quality index,

M = total length of stream.

The main problem with this index is that it does not address the situation when numerous pollutants exceed the stream standards. The user has no information as to whether 1 or 50 pollutants are a problem at a given site. The index does not consider the severity of the individual parameter violations over time. One violation occurring in

Table 2.2 Weights for Intensity Factor of PDI Index [Truett 1975]

Ecological: Inhibiting or eliminating desirable life forms.

0.1 = conditions that threaten stress on life forms (including sanitary aspects not related to verifiable instance of contagions).

0.2 = conditions that produce stress on indigenous life forms.

0.3 = conditions which reduce productivity of indigenous life forms.

0.4 = conditions that inhibit normal life processes or threaten elimination of indigenous life forms.

0.5 = conditions that eliminate one or more life forms.

Utilitarian: Reducing the economic application of the water resource.

0.1 = conditions that require costs above the norm to realize legally defined (i.e. in water quality standards) uses.

0.2 = conditions that intermittently inhibit realization of some desirable and practicable uses or necessitate use of an alternate source.

0.3 = conditions which frequently or continually prevent the realization of desired and practical uses or cause physical damage to facilities.

Aesthetic: Causing effects disagreeable to the senses.

0.1 = visually unpleasant.

0.2 = visually unpleasant with association of unpleasant tastes or odours.

a season is not distinguished from numerous violations occurring in the season. The hazards the contaminant presents at the site are not considered nor is the seriousness of different levels of the pollutant at the site.

This method is best suited for use by water quality managers who must establish priorities between various water pollution problem sites. The PDI index enables different sites or time periods to be compared on a similar basis. Rather than using a verbal rating system to describe the severity of the numerical score, the number itself is used to compare the quality between sites or over time.

None of the water quality indices described in this section are suitable for direct application into the WatQUAS Expert System. A new index, using the previously described indices as a basis, is required for WatQUAS 2.0.

2.2.2 Pollutant Loadings

Pollutant loadings are frequently used to assess the seriousness of a pollution problem and to set pollution discharge regulations. A water quality manager must be aware of the total amount of pollutant in the stream. The singular use of parameter concentration levels can often be

misleading because of the variability of the flow volume. Often, the effect of a pollutant concentration level in a small stream is a more serious problem than in a larger stream. Utilizing only a concentration measurement yields no indication of the actual quantity of pollutant passing through the stream.

Loadings to the Great Lakes are specifically monitored at 15 of the major Great Lake tributaries in Ontario. The Enhanced Tributary Monitoring Program (ETMP) was initiated in 1980 and through a flow weighted quality sampling program (more frequent sampling during high flow periods) accuracy of load estimates has been increased. Between forty and eighty water quality samples are collected every year at each ETMP site. Each sample is analyzed for eight parameters [MOE 1986];

1) total phosphorous
2) filtered reactive phosphate
3) suspended solids
4) nitrate
5) copper
6) lead
7) cadmium
8) mercury

The flow weighted sampling program is utilized to estimate loads because of the proven correlation between flow and concentration for these parameters.

Loads are commonly calculated using the simple relationship;

$$\bar{L} = \bar{Q} * \bar{C}$$

where; $\bar{L}$ = mean load,

$\bar{Q}$ = mean flow,

$\bar{C}$ = mean concentration.

This equation produces a relatively inaccurate load estimate and introduces bias. WatQUAS 1.0 utilizes this method for calculating loads.

2.2.2.1 Flow Weighted Ratio Estimators

A flow weighted ratio estimator improves the accuracy and eliminates any bias from load estimates. Many water quality sampling stations also serve as water quantity monitoring stations. If this is not the case, then the flow at a quality station may often be estimated by utilizing upstream and/or downstream flow monitoring stations. Flows are measured significantly more often than quality parameters. By assuming that flow is monitored continuously, the flow population may be used to significantly improve the parameter load estimate. The following assumptions must be recognized in order to utilize the flow weighted ratio estimator technique;

1) streamflow is monitored continuously,
2) discrete observations are available of the water quality parameter concentration,

3) flow and concentration records are distributed approximately normal.

The full population of flow is utilized to determine the mean flow (Q), this improves the estimate of the load which is based only on a sample of the concentration (C) population. The flow data utilized does not necessarily have to correspond in time with the quality data used. The flow population which is considered the most accurately measured and which corresponds to the quality data best may be used in the estimator even if the times of flow and quality measurement do not correspond. A long flow record can be utilized if the user has confidence in the data record. Use of a short record is preferable if it is deemed the most representative.

The BEALE ratio estimator is recommended by the International Joint Commission (IJC) to calculate loads for Great Lakes Tributaries [MOE 1986]. The equation for this estimator is listed below [Bodo & Unny 1983];

$$L = \bar{Q}\frac{\bar{l}}{\bar{q}}\frac{\left(1 + \frac{1}{n}\frac{s_{lq}}{\bar{l}\bar{q}}\right)}{\left(1 + \frac{1}{n}\frac{s_q^2}{\bar{q}^2}\right)}$$

$\bar{Q}$ = *mean period flow*

$\bar{l}$ = *mean sample load*

$\bar{q}$ = *mean sample flow*

n = *number of samples*

The mean square error of the load calculation is estimated using the equation;

$$S^2 = L^2\{\frac{1}{n}\left(\frac{S_q^2}{\bar{q}^2} + \frac{S_l^2}{\bar{l}^2} - 2\frac{S_{lq}}{\bar{l}\bar{q}}\right) + \frac{1}{n^2}[2\left(\frac{S_q^2}{\bar{q}^2}\right)^2 - 4\frac{S_q^2}{\bar{q}^2}\frac{S_{lq}}{\bar{l}\bar{q}} + \left(\frac{S_{lq}^2}{\bar{l}\bar{q}}\right)^2 + \frac{S_q^2}{\bar{q}^2}\frac{S_l^2}{\bar{l}^2}]\}$$

$$S_{lq} = \frac{\left(\sum_{i=1}^{n} l_i q_i - n\bar{l}\bar{q}\right)}{(n-1)} \qquad S_l^2 = \frac{\left(\sum_{i=1}^{n} l_i^2 - n\bar{l}^2\right)}{(n-1)}$$

$$S_q^2 = \frac{\left(\sum_{i=1}^{n} q_i^2 - n\bar{q}^2\right)}{(n-1)}$$

The assumption of normality which is required by ratio estimators may not always be adhered to by the two time series records being utilized. Using the mean values for flow and concentrations assumes a normal distribution of the data. Flows may vary up to five orders of magnitude in some streams and some pollutants can vary three to four orders of magnitude. This results in data being frequently skewed to the right and an over-estimation of the actual loading is the result if means are used [Bodo & Unny 1983].

The problem of skewed data can be minimized by dividing it into smaller homogeneous, approximately normal strata. A stratum is defined as a subset of the flow with the data inside a flow strata being homogeneous. Separating data into stratum associated with event and non-event flows is usually sufficient to reduce the load error.

Event flows generally contribute very little of the total pollutant load, however, they usually comprise a significant portion of the flow. Without separating the event flows and the associated concentrations, the load calculations are usually biased high. Segregating the event data eliminates this problem. The wider the range of flows and loads, the more strata that are required, two to four strata are usually adequate.

The BEALE estimator is used to calculate a load for every stratum, loads from each stratum are pooled to yield a total load for the stream. The following equation is used to aggregate the load calculations from each stratum [Bodo & Unny 1983];

$$\bar{l}_e = \frac{\sum_{j=1}^{m} N_j \bar{l}_j}{\sum_{j=1}^{m} N_j}$$

$\bar{l}_j$ = mean load of stratum j
$\bar{l}_e$ = mean period load
N_j = time in stratum j
m = number of strata defined

The mean square error of the load estimate for the entire stream is calculated using;

$$\bar{S}^2 = \frac{1}{N^2} \sum_{j=1}^{m} N_j^2 S_j^2$$

N = total time in the period
$\bar{S}^2$ = estimate of variance

The effective degrees of freedom for the load estimate are influenced strongly by the stratum with the least number of concentration samples. This is usually the stratum with the most uncertain load estimate. The following equation is used to calculate the degrees of freedom of the load estimate;

$$f = \frac{\left(\sum_{j=1}^{m} N_j^2 S_j^2\right)^2}{\sum_{j=1}^{m} \frac{N_j^4 S_j^4}{f_j}}$$

f = *effective degrees of freedom*

f_j = *degrees of freedom of stratum j*

The total load is calculated by multiplying the estimated mean load by the time within the period of interest. A flow weighted mean concentration may also be calculated by dividing the estimated load by the flow. This flow weighted mean is generally more accurate than that calculated conventionally because the problem of using non-normal skewed data is resolved by the use of homogeneous strata. The data within a stratum are approximately normal.

A confidence interval for the load estimate is calculated by multiplying the standard error for the estimate by a suitable Student t-statistic based upon the effective degrees of freedom of the estimate.

2.2.2.2 Pollutant Source Identification

The identification of pollutant loadings associated with base flows is accomplished by utilizing the strata constructed for the BEALE load calculating ratio estimator. By carefully selecting strata that isolate the low flows (base flow) from the entire flow profile, the pollutant quality samples associated with base flow may be separated. The pollutant load associated with the base flow may then be calculated.

Separating the base flow pollution is important because it allows WatQUAS to distinguish point source pollution from non-point source pollution. Non-point source pollution is primarily contributed by run-off that reaches the stream. During base flow periods there is very little run-off from the catchment, most in-stream pollution is contributed by point sources, which are always active regardless of flow condition. Information regarding pollution sources enables WatQUAS 2.0 to recognize major pollutant sources and to recommend and calculate the effectiveness of abatement measures.

3.0 WatQUAS 1.0

WatQUAS 1.0 is a prototype expert system for water quality assessment [Allen 1986]. The Mark I knowledge based system (KBS) is the initial attempt at developing an expert system for the water quality assessment of Ontario rivers. The task of conceptualizing and constructing an expert system for water quality assessment is a complicated and time consuming task. WatQUAS 1.0 is a complex and intricate series of computerized modules that link together to form an expert system. The purpose of this chapter is to briefly outline how WatQUAS 1.0 works and to examine the problems and weaknesses inherent in the system. The areas of WatQUAS 1.0 that require improvement or modification will be specifically focused on.

3.1 System Overview

This section will briefly describe WatQUAS 1.0 and also illustrate some general results. The user interacts with WatQUAS 1.0 through a series of specific commands or recognizable phrases. This type of user interface is referred to as natural language processing.

3.1.1 Data Handling

There are over 720 water quality monitoring stations in Ontario, some of which have been operating for over twenty

years. Most sites have sampling programs that consist of 10 - 30 water samples being collected yearly. The type of water quality analysis conducted for the site depends upon its classification. Figure 3.1 lists the various site classifications utilized by the Ontario Ministry of the Environment (MOE). The data, compiled and managed by the MOE are received in the form illustrated in figure 3.2. The eleven digits in the first column of the data file represents the unique location identification code for the monitoring site. The meaning of this numerical code is described below;

aabbbbcccdd

where; aa identifies the terminal basin
bbbb identifies the river basin
ccc is the station number
dd is the sample type

The second column is the date and time of sampling, the four sets of two digits in this column represent the year, the month, the date, and the hour respectively. The remaining columns represents pollutant concentrations and lab confidence codes from each sample. One purpose of WatQUAS 1.0 is the analysis of historical water quality time series records generated at these sites.

The MOE has catalogued the data from its water quality

Code	Meaning
IJC	International Joint Commission
URB	Urban
STP	Sewage Treatment Plant
SS	Special Study
RA	Regional Assessment
OPS	Other Point Source
FP	Fish Protection
IND	Industrial
AGR	Agricultural

Figure 3.1 Site Classifications

STATION DATE QUALITY MEASUREMENTS

```
                                      ALUT               ASUT
03004900102 86042211 0238800E+00ALUT 0000000E+00ASUT
03004900102 86051213 0345600E+00ALUT 0000000E+00ASUT
03004900102 86051908 0167800E+00ALUT 0000000E+00ASUT
03004900102 86052214 0245600E+00ALUT 0000000E+00ASUT
03004900102 86061407 0432400E+00ALUT               M
03004900102 86062714 0876500E+00ALUT 0000012E+00ASUT
03004900102 86081813 0564300E+00ALUT 0000102E+00ASUT
03004900102 86082913 0234100E+00ALUT 0000055E+00ASUT
03004900102 86090615 0254300E+00ALUT 0000034E+00ASUT
03004900102 86091909 0342100E+00ALUT 0000178E+00ASUT
03004900102 86092612 0879800E+00ALUT 0000211E+00ASUT
03004900102 86093011 0345800E+00ALUT 0000098E+00ASUT
03004900102 86101812 0546700E+00ALUT 0000121E+00ASUT
03004900102 86112214 0212300E+00ALUT 0000109E+00ASUT
03004900102 86121517 0412100E+00ALUT 0000067E+00ASUT
03004900102 87012913 0613200E+00ALUT 0000078E+00ASUT
03004900102 87022708 0219800E+00ALUT 0000081E+00ASUT
03004900102 87032111 0238700E+00ALUT 0000103E+00ASUT
03004900102 87040813 0298700E+00ALUT 0000099E+00ASUT
03004900102 87041519 0312400E+00ALUT 0000087E+00ASUT
03004900102 87042016 0412600E+00ALUT 0000137E+00ASUT
03004900102 87042609 0417600E+00ALUT 0000010E+00ASUT
03004900102 87050211 0567100E+00ALUT 0000019E+00ASUT
03004900102 87051213 0872100E+00ALUT 0000026E+00ASUT
03004900102 87051809 0248800E+00ALUT 0000102E+00ASUT
03004900102 87061415 0312200E+00ALUT 0000094E+00ASUT
03004900102 87071719 0514400E+00ALUT 0000072E+00ASUT
03004900102 87073011 0623110E+00ALUT 0000088E+00ASUT
03004900102 87082520 0543200E+00ALUT 0000057E+00ASUT
03004900102 87091307 0231000E+00ALUT 0000129E+00ASUT
03004900102 87102511 0290800E+00ALUT 0000078E+00ASUT
```

Figure 3.2 Water Quality Historical Record

monitoring network by the region in which each is located. Figure 3.3 shows the various regions and their location within the province. The user initially specifies to WatQUAS 1.0 the region which contains the site to be analyzed. This step directs the Expert System to the proper computer storage area, which contains the historical water quality time series data for the entire region. The user must then specify the particular site to be analyzed by the Expert System. Only one site may be analyzed at a time. WatQUAS retrieves the historical water quality time series data for the specified site, from the MOE regional data file for the subsequent numerical analysis.

3.1.2 Statistics

This section briefly outlines the statistical procedure used by WatQUAS 1.0 to numerically assess the water quality record of a pollutant. The numerical analysis is separate from the Expert System and is simply a tool to transform raw water quality records into a form which conveys the magnitude of the pollution problem to WatQUAS.

The Expert System first counts the total number of observations in the entire historical record for the parameter. Maximum and minimum concentrations are determined and the arithmetic mean, geometric mean, standard deviation and coefficient of skew are calculated.

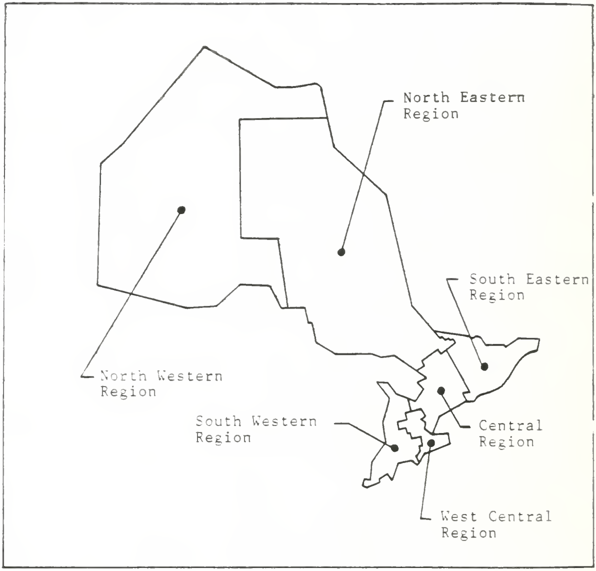


Figure 3.3 MOE Regions

Untransformed and log transformed probability distribution function (PDF) histograms are constructed and the 1st, 5th, 10th, and 50th untransformed and log transformed percentiles are calculated. The data are checked for randomness at this point to verify the assumption of independence of data required to conduct the statistical methods utilized.

Simple linear regressions are then performed on the untransformed and log transformed data to determine if a trend in the data exists. The intersect, slope and significance of both regressions are calculated. The raw data is then subjected to a randomness test in order to determine if evidence exists to support the hypothesis that seasonal trends exist in the water quality data.

The data are then broken into monthly units and the number of observations per month are counted, and the minimum, maximum, geometric mean, and standard deviation are calculated for each month for the parameter. The geometric averages are then subjected to a randomness test to determine if there is evidence of seasonality in the time series data.

A break-down of the sampling program history which shows the number of samples collected in each month for each year the site has been monitored is constructed. Monthly and

yearly total sample numbers are also tabulated for the parameter.

Figure 3.4 illustrates the results of the statistical analysis completed by WatQUAS 1.0 for one parameter. This routine is repeated for all pollutants in the MOE site record. The results from the numerical analysis are stored by WatQUAS 1.0 in a data file for future use.

3.1.3 Pollutant Correlations

Upon completion of the statistical analyses for all pollutants recorded at the site, the mean monthly concentrations are used to calculate correlation coefficients for each pair of parameters. The correlation coefficients are calculated on both categories of data.

Combinations of parameters that are significantly correlated are grouped together and a group is composed of those parameters that show significant correlation to each other. A group contains at least two parameters and any group cannot be a subset of a larger group. Figure 3.5 illustrates the correlation and grouping results from WatQUAS 1.0.

3.1.4 Violations

WatQUAS 1.0 compares all data readings for each parameter

Title: Grand River (at Dunneville) (1980-1985)
Number of Records: 576 starting: 800101 ending 851223
Parameter: PPUT
Unit of Measurement: mg/L phosphorous

Primary Statistics:

Number of Observations:	555
Maximum Observed:	1.325
Minimum Observed:	0.02
Arithmetic Mean:	0.162434
Geometric Mean:	0.130377
Standard Deviation:	0.147476
Skew:	0.0148607

PDF histogram (20 intervals of width: 0.06625)

58 256 131 40 30 13 8 6 0 1 2 1 1 2 2 0 0 0 1 2
(untransformed)

11 121 186 110 47 26 20 9 6 6 0 1 3 1 1 3 1 0 1 2
(log transformed)

	Untransformed	Log transformed
50th Percentile:	0.156179	0.14851
10th Percentile:	0.330146	0.312292
5th Percentile:	0.426803	0.412966
1rst Percentile:	0.942406	0.962606

Trends:

	Untransformed	Log transformed
Intersect:	0.186948	0.174897
Slope:	-0.00910952	-0.00724764
Significance:	-1	-1

Randomness: 0 -- Data is statistically NONRANDOM

Figure 3.4 WatQUAS 1.0 Statistical Analysis

Monthly Trends:

Month	N.Obs.	Minimum	Maximum	G.Mean	St. Dev.
Jan	15	0.052	0.375	0.111311	0.133703
Feb	35	0.058	1.2	0.206045	0.307335
Mar	87	0.02	1.325	0.15389	0.265769
Apr	115	0.056	0.93	0.164852	0.199358
May	56	0.026	0.37	0.115025	0.0754164
Jun	38	0.043	0.43	0.116013	0.097331
Jul	22	0.085	0.278	0.12703	0.0544322
Aug	24	0.052	0.2	0.124699	0.0551532
Sep	31	0.059	0.22	0.120528	0.0598079
Oct	39	0.052	0.2	0.10654	0.0421427
Nov	49	0.021	0.495	0.083192	0.0970787
Dec	44	0.027	0.445	0.104734	0.122976

Seasonality: 1 -- Sufficient evidence for seasonality

Grand River (at Dunneville) (1980-1985)
576 observations starting: 800101 ending: 851223
Parameters: PPUT NNKI NNO3FR FCMF RSP CCUT PBUT ALKT
Correlations amongst Parameters

Correlation Matrix

	PPUT	NNKI	NNO3FR	FCMF	RSP	CCUT	PBUT	ALKT
PPUT	1.00	.237	.025	.709	.790	.416	.248	-.509
NNKI	T	1.000	.926	.222	.166	.209	-.050	.229
NNO3FR	F	T	1.000	-.143	.017	-.004	-.180	.294
FCMF	T	F	F	1.000	.605	.518	.210	-.787
RSP	T	F	F	T	1.000	.399	.137	-.437
CCUT	T	T	F	T	T	1.000	.288	-.040
PBUT	T	F	T	T	T	T	1.000	-.143
ALKT	T	T	T	T	T	F	T	1.000

Figure 3.4 (Cont.) WatQUAS 1.0 Statistical Analysis

Groups -- confidence = 95

Group 1 ==>	PPUT	NNKI	CCUT	ALKT		
Group 2 ==>	PPUT	FCMF	RSP	CCUT	PBUT	ALKT
Group 3 ==>	NNKI	NN03FR	ALKT			

Log Correlation Matrix

	PPUT	NNKI	NN03FR	FCMF	RSP	CCUT	PBUT	ALKT
PPUT	1.000	.233	.053	.589	.708	.414	.255	-.439
NNKI	T	1.000	.935	.033	-.071	.221	-.048	.139
NN03FR	F	T	1.000	-.205	-.070	.028	-.153	.184
FCMF	T	F	F	1.000	.545	.641	.167	-.701
RSP	T	F	F	T	1.000	.306	.155	-.469
CCUT	T	T	F	T	T	1.000	.290	-.037
PBUT	T	F	T	F	T	T	1.000	-.124
ALKT	T	F	T	T	T	F	T	1.000

Groups -- confidence = 95

Group 1 ==>	PPUT	NNKI	CCUT		
Group 2 ==>	PPUT	FCMF	RSP	CCUT	ALKT
Group 3 ==>	PPUT	RSP	CCUT	PBUT	ALKT

Figure 3.5 WatQUAS 1.0 Grouping

to a Provincial Water Quality Objective (PWQO) in order to determine if violations of the stream standard exist. The total number of violations in the quality record are counted and the percentage of violations is calculated.

The yearly and monthly total violations are summed for each pollutant. The slope of the line constructed from the yearly violation totals is calculated in order to detect trends in the violation history. The monthly violation totals are checked for randomness in order to determine if seasonality of violations for a parameter is a problem. Figure 3.6 illustrates the violation analysis conducted by the Expert System.

3.1.5 Graphics

WatQUAS 1.0 produces three different types of graphs from the previously calculated statistics and raw time series data. Figure 3.7 shows a typical screen image of the graphics of WatQUAS 1.0. The top graph is a standard plot of the time series data versus time, lines that correspond to the mean and linearly regressed trend are superimposed on the plot.

The bottom left plot is a probability distribution function histogram of the data. The time series record is divided into 20 equal sized intervals. The number of data points

Title: Grand River (at Dunneville) (1980-1985)
576 records starting: 800101 ending: 851223
Parameter: PPUT
Summary of VIOLATIONS

555 observations -- 548 violations (98.74%)
Maximum Acceptable Concentration: 0.03 mg/L phosphorous

Average Time Between Violations: 3 days

Yearly Trend: 0 -- not significant
Seasonality: 0 -- insufficient evidence

Violations by Year:

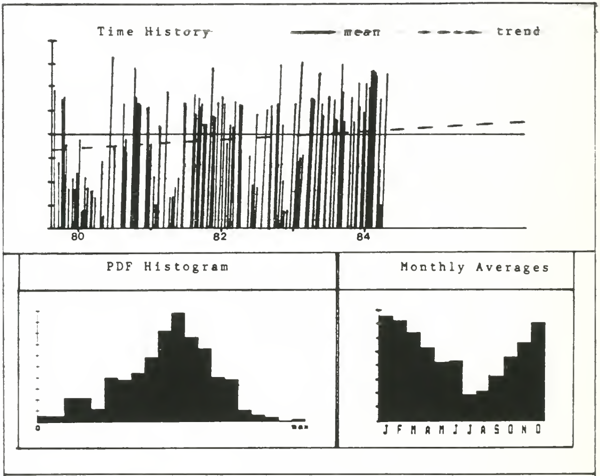
Year	Obs	Violations	(% this year)	(% total violations)
80	109	104	95.41	18.98
81	115	114	99.13	20.80
82	100	100	100.00	18.25
83	76	75	98.68	13.69
84	84	84	100.00	15.33
85	71	71	100.00	12.96

Violations by Month:

Mo	Obs	Violations	(% this month)	(% total violations)
Jan	15	15	100.00	2.74
Feb	35	35	100.00	6.39
Mar	87	86	98.85	15.69
Apr	115	115	100.00	20.99
May	56	55	98.21	10.04
Jun	38	38	100.00	6.93
Jul	22	22	100.00	4.01
Aug	24	24	100.00	4.38
Sep	31	31	100.00	5.66
Oct	39	39	100.00	7.12
Nov	49	45	91.84	8.21
Dec	44	43	97.73	7.85

Figure 3.6 WatQUAS 1.0 Violation Assessment

NNO3FR -- Grand River (at Dunneville) (1980-1985)



NNO3FR - nitrate, filt, react,
Unit of Measurement: mg/L nitrogen
Observations started: 800101 ended: 851223

STATS	TRENDS
Number of obs.:335	Intersect:2.48606
Minimum:.198	Slope:.123388 per year
Maximum:5.72	Slope:Statist. Significant
Arithmetic Mean:2.75586	Seasonality: Sufficient evidence
Geometric Mean:2.50712	
Standard Deviation:.985048	
Skew:.0208563	

Figure 3.7 WatQUAS 1.0 Graphics

occurring within an interval is plotted against the accumulated boundaries of the intervals.

The plot located on the bottom right illustrates the monthly average concentrations of the parameter. The information below the graphs is a summary of the statistics calculated from the time series data.

A geographical map of the selected region, showing the location of all the water quality monitoring stations in the region is also available to the user.

3.1.6 Water Quality Index Utilized by WatQUAS 1.0

WatQUAS 1.0 uses a water quality index to convey a measure of river water quality and the seriousness of the pollution problem. The Expert System also uses the index to recommend the strength and priority of control and abatement strategies, further investigations and water use restriction.

The present water quality index utilized by WatQUAS 1.0 has a very limited scope and application. The index can only examine and combine a maximum of nine pollutants:

1) Nitrates
2) Phosphates
3) pH
4) Temperature Deviation
5) Fecal Coliforms
6) Turbidity
7) Total Solids
8) Dissolved Oxygen
9) BOD

The water quality index overlooks a large number of pollutants, which in many cases present a more serious danger than most of the parameters listed. This index does not include heavy metals, radioactive parameters, hazardous organic contaminants and numerous "conventional" pollutants. Ideally the water quality index utilized by WatQUAS must be able to recognize any pollutant potentially found in Ontario rivers and streams.

The water quality index is a weighted product type:

$$I = \prod_{i=1}^{n} q_i^{w_i}$$

where; I = water quality index,

q_i = the individual parameter score,

w_i = the weight of the parameter,

n = the number of parameters.

The weights (w_i) are derived from curves (figures 3.8 - 3.10) which reflect 142 expert opinions regarding the effects and importance of each parameter. The "delphi" technique was utilized to poll the expert opinions.

In many cases all nine parameters are not measured at every site. If the number of pollutants recorded at the site which possess rating curves is less then 9, then "n" becomes the number of parameters used by the index and the

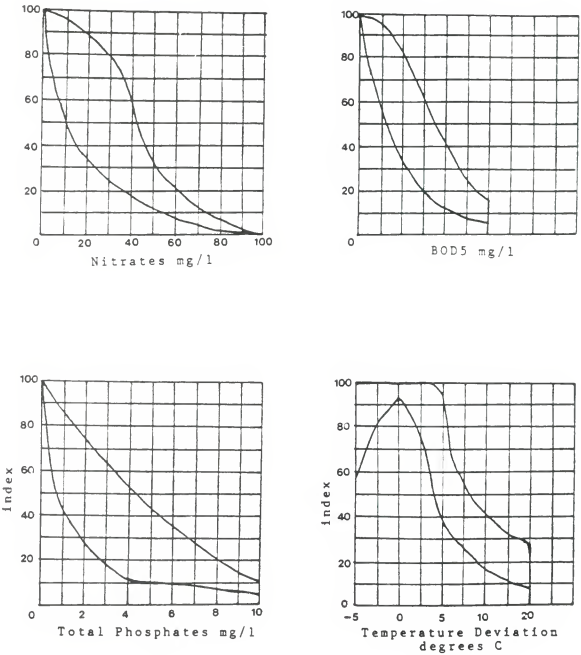


Figure 3.8 Water Quality Index Rating Curves

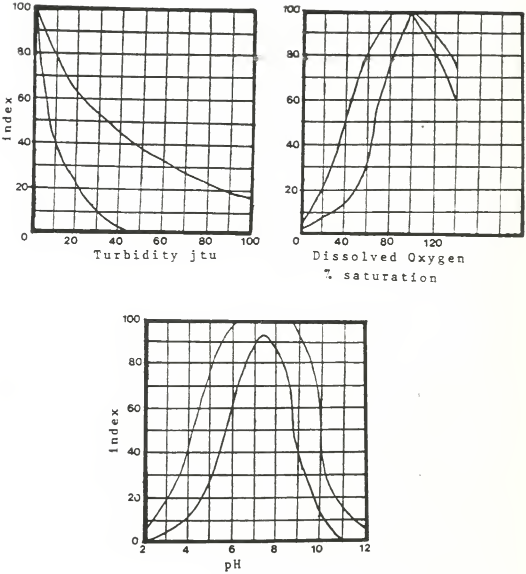
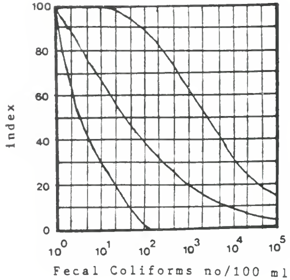


Figure 3.9 Water Quality Index Rating Curves



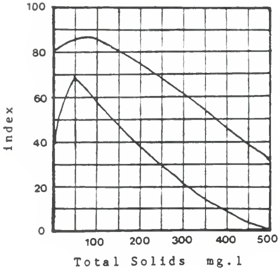


Figure 3.10 Water Quality Index Rating Curves

weights are normalized:

$$\sum_{i=1}^{n} w_i = 1$$

where; W_i = weight of the parameter,

n = the number of parameters,

This method of calculating the water quality index for a site is widely used and generally accepted. However, its inability to analyze a large number of pollutants is its biggest drawback. The contaminant scores are calculated from rating curves using the geometric mean of the water quality time record as an indicator of the pollution severity. The water quality index in WatQUAS 1.0 does not take into account stream standard violation; frequency, duration or persistence.

Figure 3.11 shows the verbal descriptions of the possible numeric values calculated by the water quality index utilized by WatQUAS 1.0. Water use at the site is not considered, sites from which water is used for drinking or for recreational purposes have the same priority as all other water uses. Similarly, environmentally sensitive sites have the same rating as non-environmentally sensitive sites.

To summarize, WatQUAS 1.0 contains a water quality index that examines only nine pollutants. The index does not

Score	Rating
0 – 25	Very Bad
25 – 50	Bad
50 – 75	Medium
75 – 90	Good
90 – 100	Excellent

Figure 3.11

Ratings for Water Quality

Index used by WatQUAS 1.0

consider individual site conditions or any type of violation analysis. One of the priorities in the development of WatQUAS 2.0 is the development of a comprehensive and robust water quality index.

3.1.7 Expert System Application

The previous sections of this review have outlined the numerical techniques utilized by WatQUAS 1.0 to analyze water quality time series data. This type of analysis is common to most water quality studies. What makes the WatQUAS Expert System unique is that it then interprets the results from the initial numerical analysis.

A conclusion regarding the overall site water quality is derived using the Water Quality Index (WQI). Various levels of quality ratings are assigned to the site depending upon the magnitude of the index. The hazard rating of the health, aquatic and economic risks are also dependent upon the WQI. An overall abatement strategy and priority is also recommended for the site which is dependent upon the ratings assigned to the various risk categories. Some examples of the various abatement recommendations are;

* Reduce human health risk;
* Reduce aquatic risk;
* Reduce economic risk.

WatQUAS 1.0 reaches conclusions regarding the water quality problems associated with each parameter that is contained in the time series record for the site and for which the Expert System contains knowledge. The expert system utilizes parameter specific information, stored in data files and the numerical results of the time series analysis, of the pollutant, to determine what the problems are and their seriousness. The geometric mean of the time series history of a pollutant is compared to the PWQO to determine if the long term pollution levels represent a problem. The significance of the slope of the linear regression of the time series record is assessed to determine if a conclusion regarding a possible trend may be reached. The time series history is inspected for missing data and gaps to determine if the data is suitable and sufficient to obtain a valid analysis. This procedure also allows WatQUAS to determine the quality of the sampling practice at the site.

A conclusion regarding the pollution problem at the site is based upon the percentage of violations, the quality trend of the parameter, the quality of data, sampling adequacy and the seriousness of the pollutant. The conclusions reached by WatQUAS 1.0 range from "no problem" to "severe problem". Various strategies such as "do nothing", "investigate cause" or "investigate STP" are recommended

depending upon the problem, the pollutant, and its sources. The priority of implementing the control strategy is also recommended.

The previous sections have contained a brief overview of the basic workings of WatQUAS 1.0. There are many finer points and details which have been overlooked in the interest of brevity. The reader is directed to [Allen 1987] for a detailed account of the development and operation of WatQUAS 1.0.

3.2 Critique of WatQUAS 1.0

WatQUAS 1.0 is a prototype expert system, it is an initial attempt at constructing a knowledge based system for water quality assessment. There are a number of problems and weaknesses with this first version which must be identified and rectified in subsequent versions of the system. However, this version serves as a learning tool and a basis for constructing subsequent versions of WatQUAS.

A standard recommendation for constructing an expert system is that after testing has been completed on the initial version it should be discarded. WatQUAS 1.0 served its purpose as a Mark I system in that testing has indicated many areas that require improvement or revision. The following critique is intended to outline these problems so

that WatQUAS 2.0 can be improved and expanded and avoid the shortcomings that are inherent in the prototype.

3.2.1 Computer Requirements

The major drawback to WatQUAS 1.0 is that it operates on a large VAX computer with the UNIX operating system. The expert system component of WatQUAS 1.0 requires a large quantity of RAM to compile and execute the OPS83 code. The site and parameter specific knowledge, stored in random data files require a large storage area. The numerical analysis of the water quality data utilizes many input/output (I/O) operations. A computer that is capable of handling and manipulating large quantities of data is required. The VAX GPX/II computer, which currently executes WatQUAS 1.0, meets these requirements.

The eventual users of this expert system, the MOE, have no capability to operate in the UNIX environment. Acquisition of the necessary hardware by the MOE, to operate in the Expert System in its current format is not foreseen in the immediate future. The result of WatQUAS being limited to the UNIX operating system is that implementation of the Expert System throughout the province, in MOE offices is not possible. This means that WatQUAS is an expert system for water quality assessment of Ontario rivers which cannot be utilized by the people who would benefit from its

assistance.

3.2.2 Natural Language Processing

WatQUAS 1.0 utilizes a simple form of natural language processing to accept commands from the user. The operator types words or short phrases to control the flow of the expert system through the water quality analysis and expert assessment. The vocabulary of WatQUAS 1.0 is very limited and unless the user knows the exact phrases or words required by the expert system, there may be difficulty operating it. A help facility is provided by WatQUAS 1.0 to assist the user. However, the natural language processing facility is cumbersome for the uninitiated person to operate.

3.2.3 Storage of Parameter Specific Knowledge

All parameter specific knowledge is stored in separate random data files. The PWQO's for all parameters are in one file, parameter source information is in a different file, while lab codes for all pollutants are stored in another file. These information files are managed outside the expert system program and accessed by the Expert System only when required. This is the best method to store knowledge because it decreases the size of the expert system and allows the operator easy access to the knowledge base.

It is a complex task for the designer and operator to manage many large data files containing numerous parameters and the extensive parameter specific knowledge that WatQUAS 1.0 requires. Many random data files introduce the problem of the user having difficulty locating specific information inside a complex array of files in order to modify or change the contents. Part of the problem can be blamed on the UNIX operating system. There is no Data Base Management System (DBMS) software available for the UNIX environment.

The heuristics of WatQUAS 1.0 contains rules used for assessing water quality problems. Parameter specific expert knowledge is contained within these heuristics. Some examples of the types of parameter specific knowledge are;

* Human health impacts
* Aesthetic impacts
* Aquatic impacts
* Socio-economic impacts
* Dissipation information
* Abatement strategies
* Maximum percentage of allowable pollutant violations permitted at a site
* Water quality index comments

This type of parameter specific information is stored in a separate rule inside the expert system component of WatQUAS 1.0 for each parameter. The number of heuristics can become excessive when many parameters are included. The modules themselves become very long when detailed and comprehensive knowledge for each parameter is added. Figure 3.12 shows the expert knowledge contained within the heuristic modules for three typical parameters. Encoding this type of knowledge within the heuristic modules inside WatQUAS 1.0, programmed in the OPS83 expert language, makes it difficult for the "non-computer expert" user to modify or change the knowledge. Every time changes are made inside a heuristic module it must be recompiled. This is an arduous and troublesome task for users who are not computer experts. Ideally much of this expert knowledge should be stored outside WatQUAS and accessed by the expert system only when required.

Some parameter specific expert information must be contained within the heuristic modules. Information that is unique or of common format to only a few parameters requires special treatment. Figure 3.13 illustrates a complex heuristic for the parameter "phosphorous". However, much knowledge, which has a similar format for all parameters can be stored outside the expert system.

```
-- alkalinity rules (ALKT)

rule ALKT_setup
(goal function=assess; object=water_quality;
status=active);
&1(parameter abbreviation=ALKT; class= ||)
-->
modify &1(class = physical;
  human_health_impact = low;
  aesthetic_impact = moderate;
  aquatic_impact = moderate;
  socio_economic_impact = high;
  dissipation = seasonal;


 -- phosphorous rules (PPUT)

 rule PPUT_setup
 (goal function=assess; object=water_quality;
 status=active);
 &1 (parameter abbreviation = PPUT; class= ||)
 -->
 modify &1(class = nutrient;
   dissipation = short;
   human_health_impact = low;
   aesthetic_impact = moderate;
   aquatic_impact = high;
   socio_economic_impact = moderate);


--lead rules (PBUT)

rule PBUT_setup
(goal function=assess; object=water_quality;
status=active);
&1(parameter abbreviation=PBUT; class=||)
-->
modify &1(class = heavy_metal;
  human_health_impact = high;
  aesthetic_impact = moderate;
  aquatic_impact = moderate;
  socio_economic_impact = low;
  dissipation = seasonal;
```

Figure 3.12 WatQUAS 1.0 Typical Rules for Three Parameters

```
rule PPUT_ask_color
 (
 
 (goal function=assess; object=env_risk;
 status=active);
 
 &1 (parameter abbreviation=PPUT;
     violation_comment<>no_problem;
     violation_comment<>mild);
 ~(parameter abbreviation=COLOR; site=here)
 -->
 
 local
    &answer : symbol;
 &answer = ||;
 write()'\n';
 while(&answer = ||) {
   write()|Is  color a problem at this site? (yes no why  ?)
   ==> |;
 
   read()&answer;
 
   if(&answer = y \/ &answer = yes) {
     modify &1(comment = a_problem);
     modify &1(strategy=rectify; strategy_object=|PPUT
     levels|);
      );
 
   if(&answer = n \/ &answer = no) {
     write() |Perhaps the phosphorous is not a problem
       |,'\n';
```

Figure 3.13 WatQUAS 1.0 Specific Rule for Lead

WatQUAS 1.0 contains expert knowledge concerning twelve parameters. Therefore, it is capable of assessing the water quality problems at a site for only twelve pollutants. These pollutants are;

1) Fecal Coliforms
2) Total Coliforms
3) Phosphorous
4) Dissolved Oxygen
5) 5-day Biochemical Oxygen Demand
6) Turbidity
7) Alkalinity
8) Lead
9) Nitrogen
10) Nitrates
11) Residual Solid Particulate
12) Copper

The number of parameters analyzed by WatQUAS 1.0 is inadequate to achieve a comprehensive water quality assessment for a river. Many conventional pollutants as well as hazardous contaminants, biological, agricultural (pesticides and herbicides) and radioactive pollutants are not analyzed or considered in the water quality assessment. The MOE Municipal and Industrial Strategy for Abatement (MISA) Effluent Monitoring Priority Pollutants List (EMPPL) contains approximately 180 hazardous contaminants which may potentially be found in the environment of Ontario. A complete expert system for water quality assessment of Ontario rivers should be capable of recognizing and assessing pollutants from any of these listed groups which may be found in Ontario.

Much of the parameter specific information contained in WatQUAS 1.0 is qualitative. The various classifications of impacts and risks are described only as low, moderate or high. The heuristics in the knowledge block, for the parameters that are present, are inadequate and superficial. There is no knowledge pertaining to the chemistry of the water pollutant, its interaction with other pollutants (synergy) or its fate in the environment. For example, the PWQO regulation for some contaminants is dependent upon the presence and ambient concentration of other parameters. The PWQO for lead is dependent upon the quantity of alkalinity present in the stream. Without alkalinity data, a meaningful assessment of stream violations for lead cannot be achieved because there is no clearly specified PWQO for lead contamination. The PWQO for cadmium is determined by a similar method, other pollutants require more complex types of analysis. WatQUAS 1.0 is not capable of assessing these types of complicated situations. Only one pollutant may be examined by the Expert System at a time.

PWQO's for pollutants which have a regulated guideline, are stored in the knowledge base in the form of a "maximum acceptable concentration". The "maximum desirable concentration" for a pollutant is also contained in the knowledge base. This number represents the maximum in-

utilized in the water quality analysis procedure. WatQUAS 1.0 does not distinguish between the two types of outliers and utilizes all data in the analysis.

Flow data utilized by WatQUAS 1.0 is in the form of monthly averages for each site. The Expert System does not contain or utilize hourly or daily time series flow records. If continuous flow data were used then a correlation of flow and pollutant concentration data could be obtained. Correlating flow and quality data would allow the Expert System to investigate possible sources and behavior of a water quality contaminant. Continuous flow data is important if the flow dependent portion of the parameter time series record is to be removed.

WatQUAS 1.0 performs a complicated statistical analysis on the water quality data. Most of the statistical procedures utilized by WatQUAS 1.0 rely on the assumptions that the data are normal and independent and that the variance is constant. Log transforming the data helps to eliminate problems associated with a skewed distribution and outliers in the time series record. However, the Expert System is assuming that the data are either normally or log normally distributed.

Water quality data frequently violate the assumptions of normality, independence and constant variance and are often

Code	Meaning
<	Actual Result < Reported Value
<=>	Approximate Result
$< N$	Non-Detected
$< R$	Detect Limit Report: Value < Limit
>	Actual Result > Reported Value
AID	Approximate Value: Insufficient Dilution
CIC	Possible Contamination Due to Improper Cap
DCP	Dangerous Constituents Present
DUP	Duplicate
M	Manually Analyzed
NSS	No Suitable Sample
RDS	Results Obtained From Diluted Sample
RVC	Value Computed From Other Results
U	Unreliable Result

Figure 3.14 MOE Comment Codes

stream concentration which does not have detrimental effects on the environment. There is no information pertaining to the specific maximum in-stream concentration limits for aquatic life, human health, recreational uses or industrial uses. Many pollutants have not been subjected to a standard regulation and do not possess a PWQO set by the province. WatQUAS 1.0 uses the 90th percentile of the pollutant time series record as the maximum allowable concentration for violation assessment if no objective is specified. This technique is not based on the chemistry, toxicity, or behavior of the contaminant. The maximum in-stream concentration of a pollutant is dependent upon the properties of the pollutant, not its time series history. WatQUAS 1.0 could easily mislead the user as to the actual situation at the assessment site.

3.2.4 Numerical Analysis

WatQUAS 1.0 does not recognize and assess the quality of the data comment codes contained with the water quality data in the MOE historical records. The meaning of these codes are listed in figure 3.14 and the understanding of their significance is vital for determining the validity and accuracy of the water quality analysis conducted at a site. Readings which are recorded as "less than the detection limit" are important because the pollutant may be

present in significant quantities. However, the concentration is not detectable using present lab analysis techniques.

The data are not examined exhaustively by the expert system to determine their quality. Problems with data quality are often encountered because of problems with the sampling technique, lab analyses and data recording. Data of poor quality may be suspected if the data are uniformly high or low for short intervals or it fluctuates erratically. There is no facility in WatQUAS 1.0 to recognize poor quality data or to manage data that requires modification or censoring. Changes in lab techniques may influence trends in the time series record. Discovering and handling problems with data quality requires judgment and knowledge for which WatQUAS 1.0 is not programmed.

Outliers in the time series data can have large effects on the water quality analysis. WatQUAS 1.0 attempts to minimize the effect of outliers by utilizing a log transformation of the data. The cause of the outliers in the time series data is not investigated by the Expert System. Some outliers are caused by poor sampling and lab analysis techniques, while other extreme sampling values represent actual in-stream pollutant levels. Data that are determined to be inaccurate or erroneous should not be

flow dependent. If these problems are encountered, WatQUAS 1.0 does not resort to alternative statistical methods but continues with the standard analysis.

The International Joint Commission (IJC) requires that loadings of certain pollutants discharging from Great Lakes Tributaries must be calculated on a yearly basis. The IJC and MOE specify that the loads must be calculated using the BEALE Ratio Estimator technique utilizing continuous flow data to improve the load estimate. WatQUAS 1.0 performs this task, however, it is accomplished by combining the average flow with the average concentration to determine the load. A Ratio Estimator is not utilized by the Expert System.

3.2.5 Overall Impression of WatQUAS 1.0

Constructing and implementing the prototype Expert System WatQUAS 1.0 was a formidable task. It contains many thousands of lines of programming and is composed of hundreds of modules. Oversights and errors are expected when a project of the magnitude of WatQUAS is attempted and completed within a short time frame. There are some logic, programming and accuracy errors in WatQUAS 1.0 that can be rectified in the second version.

Although the criticisms of WatQUAS 1.0 contained in this

chapter may seem severe, it must be remembered that the Expert System is only a prototype. Prototype systems are only the first step and are expected to change dramatically between the initial testing and the implementation of a working on-line version. Most of the problems with WatQUAS 1.0 can be rectified fairly easily by expanding the knowledge facilities and water quality assessment techniques already utilized by the Expert System.

4.0 WatQUAS 2.0

The evaluation and testing of WatQUAS 1.0 indicated that a great deal of effort would be required to construct a second version of the expert system. WatQUAS 2.0 is a step forward in the development of an expert system for the water quality assessment of Ontario rivers. The Mark II version of WatQUAS is not intended to correct all of the problems and shortfalls identified in the first version. Subsequent evaluation and testing of the second version will identify problems, limitations and shortfalls that must be rectified in future editions of WatQUAS. Many years of development and testing of this expert system will be required before a comprehensive and beneficial system can be achieved.

4.1 Development of WatQUAS 2.0

Work on the second version of WatQUAS was initiated in the fall of 1987. The testing and evaluation of the prototype had been completed and after consultation with MOE personnel a plan for the development of WatQUAS 2.0 was established in the winter of 1988. All of the individual components required to operate WatQUAS 2.0 are completed. The knowledge base has been expanded to contain information pertaining to many contaminants and the number of heuristics has been increased considerably. However, there

remains some work to be completed by a computer scientist in order to link the components and optimize the operation of the software package. A graphics capability for WatQUAS 2.0 has yet to be completed.

4.1.1 Hardware Requirements

The biggest change in the second version of WatQUAS is that it is specifically designed for execution on an IBM PC compatible computer. An IBM PC is available in most MOE departments and regional offices. Accessibility to WatQUAS by MOE personnel throughout the province will become possible.

There may be some minimum hardware configuration requirements for the computer system depending upon the capacity of RAM required to accommodate the OPS83 compiler. A large "hard-disk" will also be required for data and knowledge storage. The size of the necessary storage space, depends upon the quantity of water quality time series record that the user wishes to access. A 50 - 70 megabyte "hard-disk" is sufficient to store the historical water quality time series record for one region, the knowledge block of WatQUAS and the complete computer code.

4.1.2 Software Requirements

The water quality analysis routines and the driving

programs for the expert system component of WatQUAS 1.0 were written in C for the UNIX operating system. The second version of WatQUAS also utilizes the C language for much of its programming requirements. C - compilers for the IBM PC are readily available and are relatively compatible to the UNIX version. Minor modifications are required when translating UNIX C to IBM PC compatible C.

One reason that WatQUAS was not originally constructed to execute on a IBM PC was the lack of a powerful expert system language designed for the micro-computer. The high level expert system languages up to 1987 were available only for mainframe type computers, such as the VAX running UNIX. A version of OPS83 recently became available that is designed specifically to execute on the IBM PC. This software will enable WatQUAS 2.0 to perform fast and efficient rule tracing in the expert system component of the package.

A graphics software package is required to construct the graphs and diagrams that WatQUAS produces. The graphics in WatQUAS 1.0 were produced by the X graphics package. This software is not yet available for the IBM PC computer format. It is necessary to purchase an alternative package specifically suited for the IBM PC.

The advent of data base management system (DBMS) software

packages has permitted computer users to manipulate and manage large quantities data. The DBMS allows the user to quickly search for and retrieve the required information. Storing, editing and modifying data are accomplished simply and efficiently by using the DBMS. In WatQUAS 2.0 the water quality time series record, acquired from the MOE, is stored in a DBMS. Thus making random data files containing bulky time series data no longer necessary.

The DBMS utilized by WatQUAS 2.0 for time series record storage serves a dual purpose. First, the Expert System utilizes the DBMS to retrieve the required data for the numerical water quality analysis. WatQUAS 2.0 sends a message to the DBMS telling it what data is required and the DBMS responds by supplying the required data. The DBMS software package possesses the capability to search for data by;

* region,
* site,
* parameter,
* a specific concentration,
* date,
* comment code.

The expert system is guaranteed fast and efficient access to any type or form of data that it may require.

Instead of utilizing the entire time series record for the parameter at a site, similar to WatQUAS 1.0. The DBMS

allows WatQUAS 2.0 to access only the data within the time frame that the user desires to be analyzed. This permits the water quality within a specified time period to be analyzed. WatQUAS 2.0 also has the ability to compare the water quality data and the resulting analysis from different time periods.

The DBMS also permits the segregation and modification of data that are affected by known external factors, such as; in-stream factors, inconsistent sampling techniques or lab procedure changes. Water quality data that is of suspicious origin or quality may be eliminated from the analysis. Future work could entail developing an entire set of rules and guidelines for assessing the quality of the time series record.

WatQUAS 2.0 has access to all water quality data stored in the computer system. The simultaneous assessment of the water quality at more than one site is possible by utilizing the DBMS. This is the first step in allowing the overall quality of an entire stream or basin to be assessed. The DBMS makes it possible for the Expert System to analyze water quality over time or space.

Flow data for the water quality monitoring sites, if it is available, are also contained in the DBMS. The flow records are managed in a similar manner as the quality

data, making any portion of the flow record available to WatQUAS 2.0. The DBMS likewise permits flow data to be accessed or segregated by time, flow magnitude or station.

WatQUAS 2.0 stores data or the results of analyses in the DBMS for future use and reference. The user of the Expert System is not involved in this transfer of information as it is internal to WatQUAS. The operator accesses this data through the DBMS for viewing or modification.

The DBMS allows the user to easily locate and if necessary modify specific data records. Additional quality data and flow data are continually being supplied by the MOE and the DBMS allows the new data to be easily appended to the original record. It is possible for the operator to correct errors or oversights that have been discovered in the data record.

The second function of the DBMS software package is to manage expert knowledge. The expert knowledge is stored in the form of "memos" in the DBMS. The "memos" are accessed and read by the Expert System when they are required. The advantage of storing this expert information in the DBMS is that the knowledge can be easily modified, expanded or changed without affecting WatQUAS 2.0. The exact nature and method of storing this information is discussed in chapter 5.2.

4.1.3 User Interface with WatQUAS 2.0

Constructing a natural language processor for the expanded WatQUAS Expert System would require enormous effort and time. The number of phrases and words recognizable by WatQUAS 1.0 was barely adequate for its operation. The expanded version 2.0 would require a much larger vocabulary of words and phrases. The major emphasis in the development of WatQUAS 2.0 is to enhance the assessment of water quality in the rivers of Ontario and the expansion of the knowledge base. The luxury of natural language processing will be foregone in this version. If a comprehensive natural language processing software package becomes available in the future, it could be incorporated into a subsequent WatQUAS version. Natural language processing is an area of research separate from that of developing expert systems for water quality assessment.

WatQUAS 2.0 interfaces with the operator by utilizing a series of menus. A number, letter or word is usually sufficient to select the options required to drive the Expert System. Most software users prefer a "menu" type system to interface with the computer because of the visibility of possible options and the simplicity with which commands may be entered. Menus permit the user access to all of the features of WatQUAS 2.0. The user

can avoid consulting manuals or the help facility in order to realize the full capabilities of the Expert System.

4.1.4 Graphics Produced by WatQUAS 2.0

Water quality engineers from the Hydrology Unit of the MOE recommend "box and whisker" plots be utilized to illustrate water quality data. Figure 4.1 shows the points used to construct the box plot. The upper and lower points on the plot represent the maximum and minimum concentrations respectively. The 25th, 50th and 75th quartiles are the horizontal lines in the box and the mean is the point in the box. The geometric mean is used if the data are log transformed and the arithmetic mean is used if the data are not so transformed. The width of the box is dependent upon the number of samples being used to construct each box. Sites that are sampled frequently have wide boxes, while rarely sampled sites have narrow boxes. These plots are useful for comparing groups of data from different sites, times or parameters. Figures 4.2 - 4.4 illustrate the various forms of the "box and whisker" plots.

4.2 Operation of WatQUAS 2.0

WatQUAS 2.0 is completely menu driven, the user selects a number, letter or word from a list of options to control the operation of the Expert System. The first menu

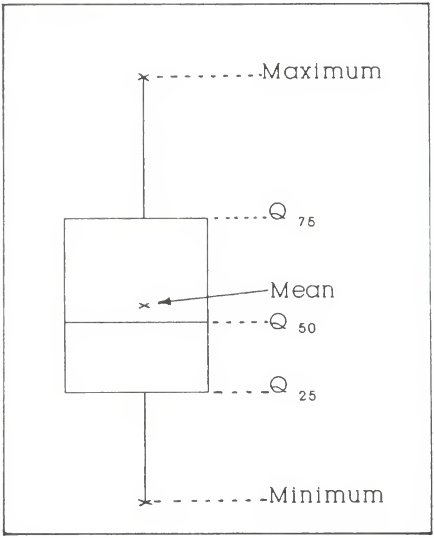


Figure 4.1 Construction of a Box and Whisker Plot

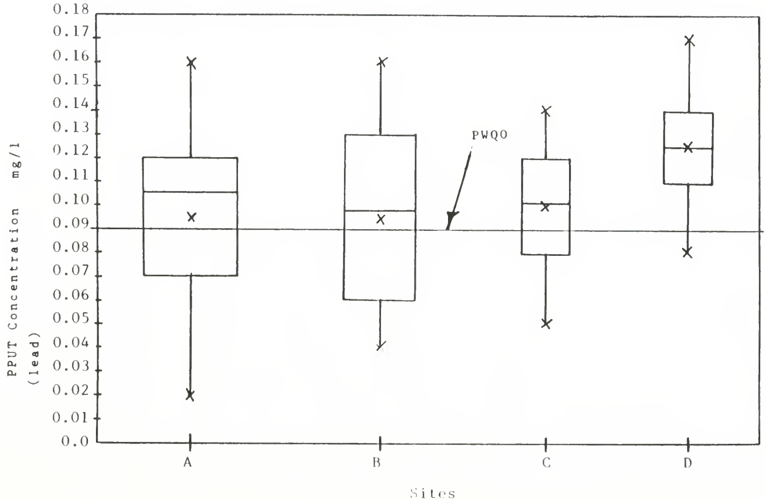


Figure 4.2 Box and Whisker Plot for Various Sites

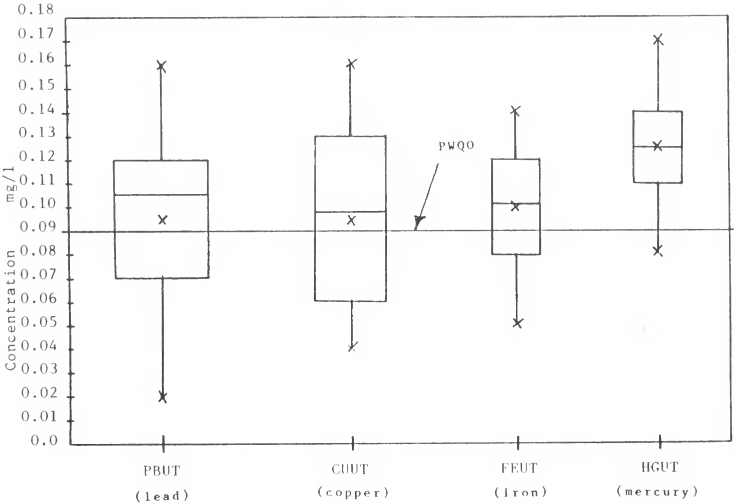


Figure 4.3 Box and Whisker Plot for Various Parameters

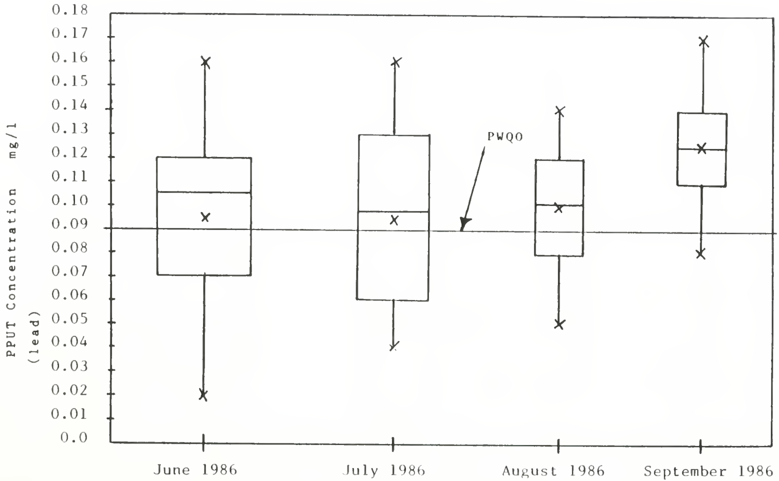


Figure 4.4 Box and Whisker Plot for Various Time Periods

confronted by the user initiates the system and allows three options;

SET
GO
QUIT

The SET option is for choosing the region and site that WatQUAS 2.0 will examine. A list of eligible regions and sites is presented to the user to assist in the selection process. The region must be selected first and the site chosen must be located within the selected region.

The GO option shifts the user directly to a general information facility. This facility permits the user to access the parameter, site and water quality situation knowledge contained by WatQUAS 2.0. The QUIT option allows the user to escape from the WatQUAS Expert System.

4.2.1 Numerical Analysis of Water Quality Data

Immediately after the region and site has been selected by the operator WatQUAS 2.0 shifts to the second menu. The main purpose of this menu is to direct the numerical analysis of the water quality data. The menu contains the following commands;

MODIFY	LIST
SHELL	SHOW
DESCRIBE	STATS
SUMMARY	IDENTIFY
HELP	GRAPH
QUIT	

The MODIFY option is for changing the previously selected region and/or site. After the numerical analysis and expert interpretation of the water quality situation at a site has been completed, the results are stored for future reference. By conducting the numerical and expert assessment at various sites, the user has the option of comparing the water quality problems, trends and analyses for the sites.

The LIST option displays the regions or sites that are available to WatQUAS 2.0 for analysis. The Expert System contains water quality and possibly flow records for these sites. By selecting SHELL the operator can execute a command from the DOS environment. SHOW displays the current region and site selections being utilized by WatQUAS 2.0. The DESCRIBE option allows the user access the general information facility.

By selecting STATS the user directs WatQUAS 2.0 to the numerical analysis module. The user selects the parameters (one or more) that require analysis and can also specify the techniques utilized to analyze the data. Section 4.3 describes the numerical analysis module in great detail.

The SUMMARY option directs the Expert System to display a summary of a selected parameter. This summary contains the results of the numerical analysis and also some general

parameter specific information that could be of interest to the operator. Figure 4.5 illustrates a typical parameter analysis summary.

The IDENTIFY option allows the user to utilize the Expert System component of WatQUAS 2.0. The Expert System component of WatQUAS is what distinguishes it from other water quality analysis packages. This module conducts an interpretation of the water quality analysis by considering the numerical results and parameter and site specific information. The STATS module must be utilized for a pollutant prior to applying the expert knowledge module. By opting to utilize the Expert System facility the user is confronted with another menu which is described in the next section.

Assistance for a problem is available by selecting the HELP option. A listing of areas for which the HELP facility is available is displayed on the screen. The operator then selects the area in which assistance is required. The GRAPH option directs WatQUAS 2.0 to the graphics module in which "whisker and box" plots, PDF's and the time series data can be displayed. The QUIT option sends the user back to the initial menu.

```
Parameter Summary for  HGFT

Full Name is  mercury filtered total

Unit of Measurement is  ug/l

class =  metal

dissipation  =  In  fresh water it is  common to be sorbed
                to  particulate  matter  and  to sediment.
                Bioaccumulation is a problem

human_health_impact =  high

aesthetic_impact =   default

aquatic_impact =  high

socio_economic_impact =  high

Chemical description: Exists primarily as Hg,Hg(I),Hg(II).
                      In  natural  water  mercury  usually
                      present  as  Hg(II).
```

Figure 4.5   Parameter Summary Produced by WatQUAS

4.2.2 Expert Assessment of Water Quality

The expert system component of WatQUAS 2.0 is capable of assessing many different water quality concerns. The user selects an area from the following list and WatQUAS 2.0 searches for knowledge that may solve the problem or enable it to reach a conclusion regarding the specific situation. Chapter 5 describes the knowledge engineering and application of the expert knowledge. A brief description of the various options available to the user in the expert knowledge module is presented in this section. The selections that the user can choose to access the expert assessment are;

* SITESUM
* PROBLEMS
* ABATEMENT
* FATE
* HELP
* QUIT
* PARSUM
* TOXICITIES
* PLANNING
* DESCRIBE
* GRAPH

The SITESUM option produces a summary of the expert assessment of the overall numerical water quality analysis. The water quality index, violation assessment, trend analysis, and statistical analysis are all used by WatQUAS 2.0 to draw conclusions regarding the water quality at a site. The Expert System conducts an interpretation of such things as;

* The pollution trends,

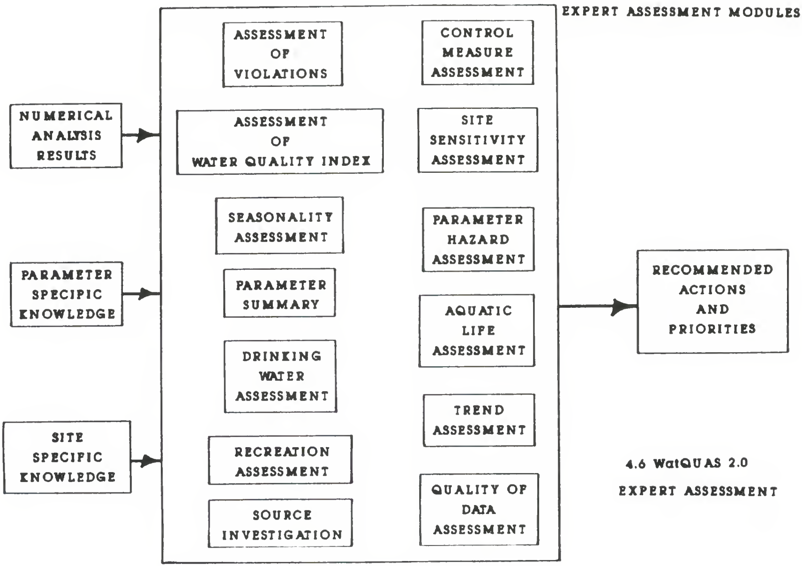
* The significance of the water quality index,
* Specific site impacts from the pollution,
* The quality of the data and sampling efficiency,
* The seriousness of the pollution problem.

PARSUM produces a summary of the expert assessment of the problems associated with a particular pollutant at a site. The operator selects a valid pollutant from a list of parameters from which the numerical analysis has been conducted. WatQUAS 2.0 then uses the expert knowledge to interpret the significance of the numerical results. The areas in which a parameter specific expert interpretation is conducted are;

* The seriousness of the problem the contaminant presents,
* Insight into the likely sources of the pollutant,
* The specific effects of the pollutant at the site.

Figure 4.6 illustrates the procedure utilized by WatQUAS 2.0 for completing the expert assessment for a typical parameter.

By selecting the PROBLEMS module, the user directs the Expert System to examine site specific problems. The site specific information is compared to contaminant levels, violation history and pollution trends in order to determine problems at a site. Potential problems are



4.6 WatQUAS 2.0
EXPERT ASSESSMENT

determined by examining pollution trends and site sensitivity. The site knowledge contains such details as; designation as a fish spawning area, recreational usage of the water or drinking water usage. A priority rating is assigned to the problem and a suggestion of the immediate action required at the site is provided.

The TOXICITIES options directs WatQUAS 2.0 to examine the health related problems the contaminants present at a site. Toxicity and health hazard ratings are used to determine the specific effects that a contaminant can have on human, aquatic and plant life. All of the pollutants that present a health hazard are assessed.

By selecting ABATEMENT the user directs WatQUAS 2.0 to produce a series of control measures and abatement actions aimed at rectifying the pollutant situation at a site. The priority with which the recommended control measures should be implemented is presented. The priority is dependent upon pollutant levels, site sensitivity and the toxicity of the contaminant.

The PLANNING selection allows various planning strategies for the river and surrounding area to be investigated. Additional pollutant loadings from proposed development, industrial or agricultural sources are compared to present pollutant levels to determine the impacts and effects of

increased pollution levels. The minimum seven day consecutive low flow with a 20 year return period (7 LQ 20) is utilized by WatQUAS 2.0 to determine the maximum concentrations from point source pollution based on minimum dilution requirements. This module requires further development to include planning strategies in the knowledge base.

By selecting the FATE module, the user is presented with the fate and dissipation information of a contaminant. The "half life" (t 1/2) of the chemical in the aquatic environment, the accumulation (biological and environmental) and dissipation of the contaminant in the stream are presented. The effects that other pollutants in the stream may have on the fate of the contaminant are also interpreted.

The DESCRIBE option sends the user to the general information facility. The HELP selection directs the user to information on the operation of WatQUAS 2.0. The GRAPH option shifts the operator to the graphing module and the QUIT selection ejects the user from the expert system component of WatQUAS 2.0.

Figure 4.7 illustrates a flow chart describing the operation of WatQUAS 2.0. The expert assessment modules of figure 4.6 are contained in the "expert assessment" box of

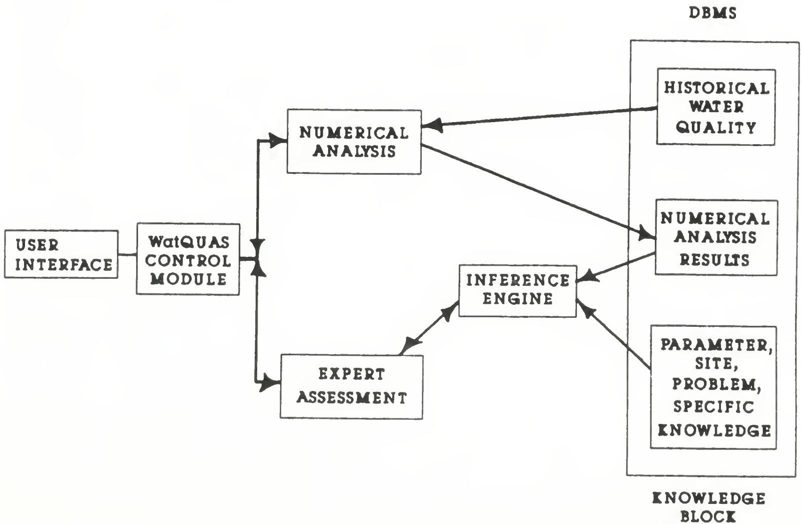


Figure 4.7 WatQUAS 2.0 Flow Chart

figure 4.7. The water quality analyses operation and the expert system component interaction of WatQUAS 2.0 are described by the flow chart.

4.3 Water Quality Assessment Methods Utilized by WatQUAS 2.0

The WatQUAS Expert System is flexible in that it can be programmed to utilize the results from various statistical analyses or water quality assessment techniques. The numerical analyses of the water quality data are conducted by WatQUAS 2.0 through a series of independent modules. Each module contains a computerized water quality assessment technique. The specific techniques employed and the order of analysis are regulated by the expert component of the system and the user. Alternative water quality assessment techniques or statistical methods can be easily integrated into WatQUAS at the discretion of the user. This section outlines the statistical methods and water quality assessment techniques utilized by WatQUAS 2.0.

4.3.1 Statistical Analysis

The modules containing the statistical analysis techniques used by the Expert System are separate "stand alone" components. Different statistical assessment packages can be incorporated into WatQUAS easily. Individual operators may have their own preference of the specific methods to be

utilized for conducting a numerical analysis. WatQUAS 2.0 can be customized to perform the type of analysis which best suits the applications of the user. The numerical analysis techniques do not affect the operation of the expert component of the system. Under the direction of MOE personnel [Bodo 1988] and [Ward & Loftis 1986] a thorough and robust statistical analysis routine was developed for WatQUAS 2.0.

Assumptions on the part of the user, of data normality, independence and constant variance have been the downfall of many statistical analyses. The major emphasis of the statistical package employed by WatQUAS 2.0 is to avoid relying upon unverified assumptions with regards to the water quality data.

Statistical assessment techniques can be categorized into two groups; parametric and non-parametric methods. Parametric statistical techniques assume that the the data adheres to a specified distribution that is described by at most three parameters. Some common parametric distributions often used in water quality analyses are; normal, log normal and Gaussian. Parametric statistical methods are usually comprised of the calculation of means, standard deviations, coefficients of skew, t-tests, linear regression, etc.. The problem with utilizing parametric

techniques for water quality analysis is that water quality data is often difficult to categorize into a prespecified distribution. Water quality data are often skewed, are not independent and do not possess a constant variance. If parametric methods are employed without verifying the distribution of the data or the validity of the assumptions the analysis is based upon, then the results may be unreliable. Parametric techniques do not allow comment codes (such as > or <) to be utilized, only exact numbers are acceptable.

Non-parametric statistical techniques are not reliant upon the distribution of the data. For water quality data analysis, non-parametric techniques are usually comprised of a rank - order analysis of the data. The data are subjected to calculation of quartiles, quantiles, the median and maximum and minimum values. The major advantage to this type of assessment is that no assumptions pertaining to the distribution shape, independence of the data or constant variance are required. Non-parametric techniques are capable of assessing comment codes.

WatQUAS 2.0 completes a statistical analysis for parameters present in the MOE water quality times series record for a site. The user has the option of selecting a specific time period from the water quality record for analysis. The

Expert System analyzes the entire record of the water quality data if time period is requested by the user. The quality record of a pollutant may be divided into more than one time period for analysis by WatQUAS 2.0. After the analysis of each parameter and time period is completed, the results are stored in the DBMS for future use in the expert assessment and for graphing.

4.3.1.1 Data Inspection

WatQUAS 2.0 inspects the water quality time series record for missing data or gaps in the sampling history. This information is utilized by the expert component of WatQUAS 2.0 to reach a conclusion regarding the quality and adequacy of the sampling program at the site. The rules and methods used to determine the quality of data and sampling practice are similar to the techniques employed by WatQUAS 1.0.

4.3.1.2 Non-parametric Statistics

The first step completed by WatQUAS 2.0 in the statistical analysis is to conduct a rank - ordering procedure on the water quality data. The number of values in the time series data record are counted and the percentage of readings that are recorded as greater than the detection level are calculated as a function of the total number of samples.

The next step in the analysis is to determine the maximum and minimum data readings. If the quotient of the maximum concentration over the minimum concentration is greater than 20 than a log transformation of the data for that parameter is performed. WatQUAS 2.0 then calculates the 25th, 50th, and 75th quartiles of the record. The quartiles are calculated by using the formula;

$$i = p * (n+1)$$

where n = total number of data in the time series,

p = the desired quartile ie. 75th (i = 75),

i = the rank - order position of the required data point.

For example, with the data sorted in ascending order, and n = 47, the 75th quartile would be;

$$.75 * (47+1) = 36 = i$$

The water quality sample at position 36 would be the 75th quartile. If the calculated i had not been an integer, but rather a decimal fraction, then a linear interpolation would be used to calculate the desired quartile. For example, if i = 36.4, and the data corresponding to ranks 36 and 37 are .61 and .64 respectively. The 75th quartile would be calculated to be;

$$Q75 = .61 + [(36.4 - 36) / (37 - 36) * (.64 - .61)]$$

$$= .622$$

The advantages of using quartiles is that they indicate the positioning and distribution of the data without requiring WatQUAS 2.0 to make any assumptions regarding the distribution shape or properties of the water quality record.

4.3.1.3 Outliers in the Data

A major concern of WatQUAS 2.0 in the statistical analysis is the presence of outliers in the data. A simple technique is utilized by WatQUAS 2.0 to identify outliers. The interquartile range (IQR) is calculated by subtracting the 25th quartile from the 75th quartile;

$$IQR = Q75 - Q25$$

The IQR indicates the size of the span from the the central 50 percent of the data.

One method of detecting outliers is to set up "fences" in the data set. Inner and outer reference points are calculated and form the fences. The data values outside the outer points are defined as "far out" and data between the inner and outer points are defined as being "out". The inner reference points or "inner fence" is calculated by;

$$Q75 + S \quad \text{and} \quad Q25 + S$$

$$\text{where} \quad S = 1.5 * (Q75 - Q25) = 1.5 * IQR$$

The outer fence is calculated by;

Q75 + 2S and Q25 + 2S

WatQUAS 2.0 segregates the "far out" and "out" data points and may delete them from subsequent use at the discretion of the operator. The operator must decide if the outliers should be deleted from the data set or utilized in the analysis. The "out" and/or "far out" outliers may be deleted by the user depending upon their acceptability as realistic water quality data values. Ideally, the Expert System should make the decision regarding the inclusion or deletion of outliers. However, the rules required to accomplish this decision are not clear and very complicated. Future versions of WatQUAS may incorporate an outlier evaluation module in the expert component of the system.

If the operator chooses to delete any outliers then the entire analysis process is recalculated omitting the deleted data points. The recalculated "far out" and "out" data values are presented and the user may choose to delete the new outliers. This process continues to recycle until the operator is satisfied with the contents of the time series record to be used in the subsequent analysis.

4.3.1.4 Determining the Distribution of the Data

The IQR is utilized to calculate a "quartile coefficient of skew" (Cs);

$$Cs = \frac{Q75 - (2 * Q50) + Q25}{IQR}$$

The skew coefficient indicates the skewness of the data distribution contained within the Interquartile Range of data. A water quality time series with no skew in the central 50 percent of the data would have a Cs = 0.

The major emphasis of the analysis has been to quantify the form of the data and to deal with outliers. All of the procedures conducted have been robust, non-parametric techniques. The results of this analysis are utilized by WatQUAS 2.0 for the construction of "Box and Whisker plots". The Expert System and the operator can interpret many things from these plots. Such as;

* Trends in the water quality data,
* Skewness in the quality record,
* Seriousness of the pollution problem.

If WatQUAS 2.0 is satisfied that the coefficient of skew is within an acceptable range and that the box plots appear relatively normal then the data is assumed to be

approximately normal or log-normal. The means and standard deviations of the data are then calculated. Arithmetic means are used if the data is not transformed and geometric means are used for log transformed data. Figure 4.8 illustrates a summary of the analysis completed for the water quality record of one parameter.

The data for each parameter is grouped by month, and year. The sample size, three quartiles, maximum and minimum concentrations, mean, standard deviation, IQR, and coefficient of skew are calculated for each time grouping. The results of these analyses are catalogued and stored by WatQUAS 2.0 for future reference and graphing.

4.3.1.5 Normal and Log-Normal Distributions

If WatQUAS 2.0 has determined the data to be approximately normal or log normal and the operator agrees with this assessment, then a variety of parametric statistical techniques are employed. This part of the assessment utilizes a modified version of the statistical analysis module constructed for WatQUAS 1.0. In the correlation module, parameters are correlated and divided into groups if significant levels of correlation exists in their records.

Autocorrelation and seasonality analyses are conducted on

WatQUAS 2.0 Non-Parametric Parameter Analysis

```
Site = Grand (at Dunneville)

Parameter = PBUT
Full Name = lead unfiltered total
Time Period of Analysis = 860304 - 871022
n = 35
% > than Detection Limit = 100%
Max = .090 mg/l
Min = .003 mg/l

Max/Min > 20
Data has been Log Transformed

Q25 = .013 mg/l
Q50 = .051 mg/l
Q75 = .074 mg/l

Coefficient of Skew = .20

Geometric Mean = .047 mg/l

Standard Deviation = .021
```

Figure 4.8 Standard Format of a Non-Parametric Pollutant Analysis Conducted by WatQUAS 2.0

the water quality record to determine the independence of the data and the seasonal trends of the water quality data. These techniques in WatQUAS 2.0 are similar to the methods employed by the prototype version. The reader is directed to [Allen 1987] for a comprehensive description of these statistical techniques being utilized by the WatQUAS Expert System.

For water quality trend assessment a simple linear regression, similar to WatQUAS 1.0, is utilized. An advanced technique for determining long term water quality trends that does not require unverified assumptions is in the final stages of development by the MOE. This algorithm may replace the present linear regression module in WatQUAS upon its final completion and testing.

The results from the parametric statistical assessment produced by WatQUAS 2.0 are very similar to WatQUAS 1.0. Section 3.1.2 describes the results from this type of assessment. If WatQUAS 2.0 is not satisfied with the quality of the data then these techniques are not be employed, unless directed by the operator.

4.3.2 Violation Assessment

If Provincial Water Quality Objectives are not specified for a contaminant then standards from other sources, such

as the Canadian Water Quality Guidelines [CWQG] are utilized. This is one area of domain knowledge which must be continually updated as new guidelines are imposed and more contaminants are subjected to regulation.

If a water quality standard for a hazardous contaminant can not be obtained from a reliable source and the pollutant is toxic and contained in the MOE Effluent Monitoring Priority Pollutant List, then detection of the contaminant constitutes a violation. WatQUAS 2.0 informs the operator very clearly if this technique for assessing violations is invoked. The number of violations occurring in the data set are reported in the form of the percentage of samples which exceeded the standard. This version of the Expert System employs a similar violation tabulation method as WatQUAS 1.0 (section 3.1.2).

4.3.3 Cumulative Distribution Functions

Cumulative distribution functions (CDF's) are constructed from the historical time series record of a pollutant. They are used to determine the probability that a sample will be in violation of the water quality standard [Loftis and Ward 1981]. If the standard is represented as an upper limit then the probability of exceedance is;

$$P[X > X_S] = 1 - F(X_S)$$

where; P[ ] = the probability of the event inside the brackets,

X = the value of a water quality sample,

X_s = the stream standard,

$F(X_s)$ = the CDF of X .

The estimated number of violations in a water quality record is;

$$\text{Number of Violations} = [1 - F(X_s)] * N$$

where; N = the number of samples in the water quality record.

Unless there is continuous quality monitoring at the site, a CDF can not accurately predict the fraction of time that a pollutant will violate the stream standard. An example of a CDF is illustrated in figure 4.9.

4.3.3.1 Procedure for a Non-Parametric Distribution of a CDF

If WatQUAS 2.0 concludes that the quality record of a pollutant does not adhere to a known distribution then a non-parametric technique is used to calculate the value of the CDF for any sample within the quality record. The Expert System has previously ranked the water quality record in ascending order, the value of the CDF for any point is given by;

$$F(X) = \frac{M}{N}$$

SAMPLE CUMULATIVE DISTRIBUTION FUNCTION
FOR LEAD

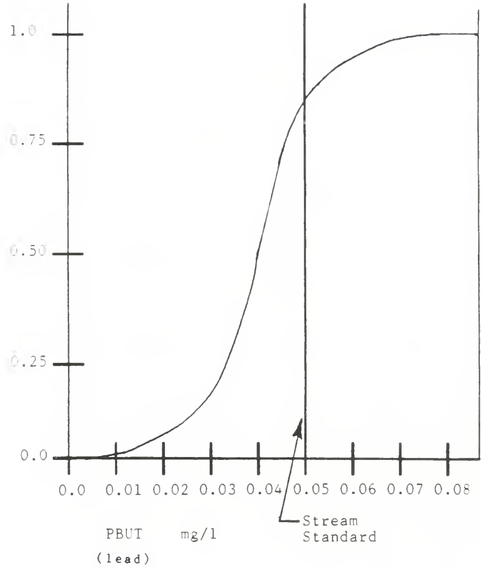


Figure 4.9 Sample Cumulative Distribution Function

where; F(X) = the value of the CDF at point X,

M = the number of observations less than or equal to X,

N = the total number of observations.

4.3.3.2 Procedure for a Parametric Distribution of a CDF

When WatQUAS 2.0 has determined that the quality record adheres to a normal or log normal distribution a more complicated procedure than the non-parametric technique is utilized. An estimate of the value of the CDF for any point is given by;

$$F_n(X_0) = \theta\left(\frac{X_0 - \bar{x}}{S}\right)$$

$F_n(X_0)$ = *the sample estimate of* $F(x_0)$

$\theta()$ = *the standard normal CDF*

x_0 = *some fixed point*

WatQUAS 2.0 utilizes the value of F(X) to determine the probability that a future water quality sample will be a violation of the stream standard.

4.4 Development of a New Water Quality Index for WatQUAS 2.0

The water quality index used by WatQUAS 1.0 is inadequate and requires major revisions. None of the indices in the technical literature were found to be entirely suitable, in their original form. The index which examined the most pollutants, contained only 72 rating curves. This is not nearly sufficient for the expert system to draw a conclusion regarding the overall water quality. There are many other problems with the various indices which have been described previously in section 3.1.6. WatQUAS 2.0 requires a water quality index constructed specifically for Expert System application, transferable to all Ontario rivers and streams and capable of considering any pollutant potentially found in the province.

The new index for WatQUAS 2.0 produces two numbers to express overall water quality. The first number represents an index similar to the PDI index discussed previously in section 2.2.3. This index accounts for the conditions at the site. The second index is a parameter specific index, which examines and aggregates the quality of individual pollutants.

4.4.1 General Site Index

The PDI index is modified to represent the quality at a

single site instead of over the entire stream. Both prevalence (P) and stream length (M) are set equal to 1.0, in order to eliminate them from the index equation. It would not be possible to determine the length of pollution in streams using the typical water quality historical record available in Ontario. Most data are collected at a single established sampling location. Future versions of WatQUAS may examine the entire stream quality, the original PDI index would then be applicable . The form of the modified PDI index is:

$$V = D * I$$

V = water quality index
The weights for D are unchanged
The weights for I are listed in table 4.1.

This index enables the user to compare the magnitude of pollution problems between various sites. It also allows the user to examine the water quality trends of the site over time. For the water quality manager it permits him/her to assess the usefulness and efficiency of implemented control programs.

4.4.2 Parameter Specific Index

The second index proposed for WatQUAS is parameter specific, the effects of each pollutant are considered in

Table 4.1 Weights for Intensity Factor of PDI Index for WatQUAS 2.0 [Truett 1975]

Ecological: Inhibiting or eliminating desirable life forms.

0.1 = conditions that threaten stress on life forms (including sanitary aspects not related to verifiable instance of contagions).

0.2 = conditions that produce stress on indigenous life forms.

0.3 = conditions which reduce productivity of indigenous life forms.

0.4 = conditions that inhibit normal life processes or threaten elimination of indigenous life forms.

0.5 = conditions that eliminate one or more life forms.

Utilitarian: Reducing the economic application of the water resource.

0.1 = conditions that require costs above the norm to realize legally defined (i.e. in water quality standards) uses.

0.2 = conditions that intermittently inhibit realization of some desirable and practicable uses or necessitate use of an alternate source.

0.3 = conditions which frequently or continually prevent the realization of desired and practical uses or cause physical damage to facilities.

Aesthetic: Causing effects disagreeable to the senses.

0.1 = visually unpleasant.

0.2 = visually unpleasant with association of unpleasant tastes or odours.

calculating the index. Since the task of constructing rating curves for all of the pollutants potentially found in Ontario rivers would be prohibitive , rating curves are not utilized for any parameter.

Many hazardous contaminants do not have established PWQO's. Their effect upon humans and aquatic life forms is often not known or fully understood. This makes it impossible to judge the severity of the "in stream" concentration of the pollutant. The number of pollutants potentially found in the environment and the lack of hard information concerning each pollutant makes it unlikely that PWQO's will be set for many pollutants in the near future. The magnitude of the concentration will not be considered in this index. The detection alone of a toxic substance, at a site, is sufficient for calculating the water quality index.

The knowledge block of WatQUAS contains detailed parameter specific information such as the type illustrated in figure 4.10. A parameter is categorized as either toxic or non-toxic. The information from the sub-section "IMPACTS" of figure 4.10 is used to compute the parameter specific index. The four impacts are each assigned an importance weight:

human health impact = 0.4

aquatic impacts = 0.4

Human Health Impact = High
Aesthetic Impact = default
Aquatic Impact = High
Socio-Economic Impact = High

IMPACTS FROM LEAD

Figure 4.10 Parameter Impact Knowledge

socio-economic impacts = 0.1

aesthetic impacts = 0.1

total = 1.0

These weights are assigned on the basis of priorities. Human health and aquatic life are assigned the highest priority, while socio-economic and aesthetics are assigned lower priorities. The weights are arbitrary and may be changed by the user depending upon their preference.

The knowledge base contains various descriptions (default, no impact, low, moderate, or high) of each impact. Default indicates that no information exists in the knowledge base concerning the impact of the pollutant. The water quality index assigns a numerical score for each level of description:

no impact = null

low = 3

moderate = 6

high = 10

default (non-toxic pollutant) = null

default (toxic pollutant) = 8

Assigning null to a "no impact" rating attempts to eliminate the problem of eclipsing. Pollutants which present no danger and possess low scores cannot mask or

hide the more hazardous contaminants. Only pollutants which present some danger are included in this water quality index.

If no information is known about an impact (default) of a non-toxic pollutant then the score of the impact is considered to be null and the impact is eliminated from further consideration. For a toxic pollutant the arbitrary default score is eight. It would be unreasonable to assign a score of zero to the impact of a hazardous contaminant. It is also too conservative to assign a score of ten (the highest) to an impact when we have incomplete information. The score of eight represents a compromise that scores the uncertain impact sufficiently high without being too conservative. These scores may also be changed by the user at their discretion. The weights are normalized for the total number of impacts used. The various impacts are aggregated using the weighted product method:

$$I = \prod_{i=1}^{n} q_i^{w_i}$$

TOXIC POLLUTANT
Example: Human Health Impact = 6
Aquatic Impact = default
Socio-Economic Impact = 10
Aesthetic Impact = 10

$$I = 6^{.4} * 8^{.4} * 10^{.1} * 10^{.1} = 7.5$$

All pollutants are weighted evenly when aggregated, the scores from the impacts is sufficient to distinguish the seriousness of the parameters. The unweighted product method is used to combine the individual contaminants;

$$I = \left(\prod_{i=1}^{n} q_i \right)^{\frac{1}{n}}$$

Example:

lead = 9
fecal coliforms = 8
carbon tetrachloride = 8.5
polychlorinated biphenyls = 10
phosphates = 7.5
nitrates = 4

$$I = (9 * 8 * 8.5 * 10 * 7.5 * 4)^{1/6} = 7.5$$

The water quality index at the site is 7.5. WatQUAS 2.0 then utilizes a scoring system to convert the numerical value into words. Figure 4.11 illustrates the various descriptions of the WQI levels.

WatQUAS 2.0 lists the individual pollutant scores (numerical and verbal) and the overall water quality index (numerical and verbal) for the specified time period.

The new water quality index used by WatQUAS produces two numbers. A modified PDI index is used to compare the seriousness of the effects of pollution problems at different sites or time periods and for a violation analysis. A parameter specific index is used to judge the

Score	Rating
9 – 10	Very Bad Situation; Extreme Problem
7.5 – 9	Bad Situation; Serious Problem
5 – 7.5	Moderate Problem
2.5 – 5	Some Concern
0 – 2.5	No Immediate Concern

Figure 4.11

Ratings for Water Quality

Index used by WatQUAS 2.0

overall water quality problem at a particular site or time period. Rating curves are not utilized by this index, instead, the impacts of each pollutant upon various sectors is used to score the pollution problem.

This new water quality index requires further refinement and testing, a comprehensive and robust water quality index will be the result, specifically suited for the Expert System application.

4.5 Pollutant Loadings

The original computer algorithm to calculate loads using the BEALE ratio estimator was developed by the IJC Great Lakes Section, Windsor office. Subsequent modifications have resulted in the algorithm being reprogrammed in BASIC and adopted specifically to analyze pollutant loadings at ETMP sites. The time period for a loading calculation is either one calendar or one water year (October to September) in this program.

In order to incorporate the BEALE estimator into WatQUAS, it had to be rewritten in "C". The load calculating algorithm, obtained from Ontario Ministry of the Environment, River Systems Unit personnel, was modified when it was reprogrammed to make it much more versatile. The load calculating program used by WatQUAS 2.0 is able

to:

* calculate the load for any length of time period for which concentration data exists,
* calculate confidence intervals for the estimated loading using the Students t statistic,
* calculate a flow weighted mean concentration.

WatQUAS 2.0 arbitrarily segregates the flow record into nine strata, customized flow strata selection is the responsibility of the user. No reliable methodology was discovered that could segregate flows into proper strata for all possible flow regimes and streams. Much work would be required to develop an expert system module that could determine the optimum number and limits of the required flow strata. The experience and intuition of the user is considered to be more reliable than a computerized method at this point in time.

The expert system does assist the user in selecting the proper strata by displaying the flow history of the stream. The system also maintains the constraint that all strata must contain at least two concentration records. Loads calculated for strata using less than two concentration records have a very large variance and add to the uncertainty of the entire load estimate.

The results of using the BEALE ratio estimator to calculate

phosphorous loads for the Grand River at Dunneville for the 1985 water year are given in figure 4.12.

4.5.1 Pollutant Load Reduction

The best method to judge the effectiveness of pollution abatement strategies is to examine the overall pollutant load reductions over a corresponding period of time in the stream. Water quality pollutants are contributed from either point sources or non-point sources. Specific pollution control strategies are usually directed towards one of these sources. Proper use of the BEALE ratio estimator for calculating pollutant loadings allows WatQUAS 2.0 to distinguish between point and non-point source pollutants.

4.5.1.1 Point Source Pollutant Reductions

As described earlier, point source pollutant loadings are not flow dependent and usually remain constant regardless of the flow or season. Typical point source dischargers are industrial manufacturers, food processing factories, municipal waste treatment outfalls, etc.. The location of point sources can be identified and the discharge monitored accurately to measure the effectiveness of control options. WatQUAS 2.0 requires the user to enter a percent reduction in pollutant load contribution for a specific pollutant and

```
SUMMARY FOR THE 4 STRATA;
THE ESTIMATED MEAN DAILY LOADING IS 2047.646   kg/day
                           + or - 148.6953   kg/day
                                  7.26   %

THE ESTIMATED LOADING FOR THE DESIGNATED
                 TIME SPAN IS:  747.4  tonnes

THESE ESTIMATES ARE BASED ON 1.52 EFFECTIVE DEGREES OF
FREEDOM
```

Figure 4.12 Results of Pollutant Load Analysis

an estimated cost for the control action. The expert system calculates the reduction in point source pollutant load. It then combines this with the non-point source load to determine an overall percentage reduction in pollutant loading in the stream due to the point source control action. The marginal cost of the specific control option is calculated and this can be used to compare various water quality management options. A record of the effectiveness and cost of the control strategy (entered by the user) is retained by WatQUAS 2.0 for future reference. Figure 4.13 illustrates this procedure for examining point source pollutant loadings.

4.5.1.2 Non-Point Source Pollutant Reductions

Calculating non-point source load reductions is more complicated than point source loadings. The non-point source load varies in each flow stratum except base flow, where it is assumed to be zero. WatQUAS 2.0 sums the individual non-point source loads from each stratum to achieve a total non-point source load.

For non-point sources the user may alternatively enter a percentage reduction in pollutant loading with an estimated cost. WatQUAS 2.0 determines the overall effect of the reduction and compares it with other options. In this way, WatQUAS can advise the user of the optimum method to reduce

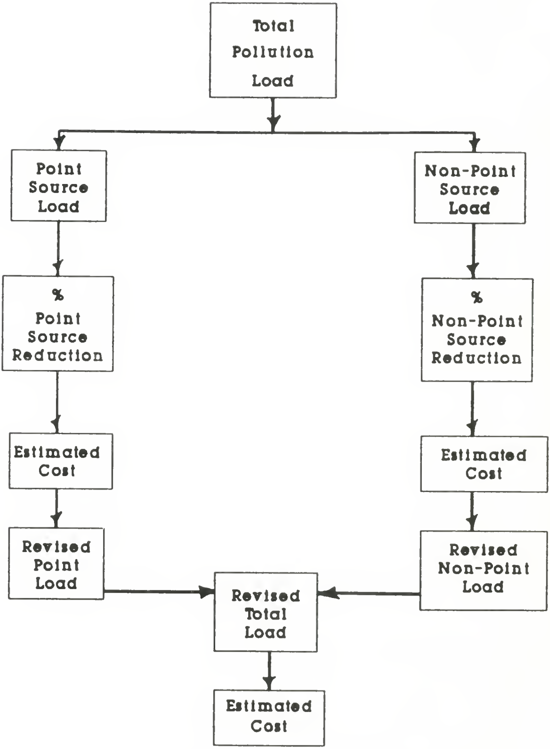


Figure 4.13 Pollutant Load Reductions

pollutant loadings. By separating and identifying the pollutants sources, WatQUAS 2.0 enables water quality problems to be examined and control measures suggested.

The knowledge block of WatQUAS contains general control suggestions and the estimated percentage effectiveness of the control measure for each pollutant. The estimated cost of each control measure is the responsibility of the user to furnish to WatQUAS 2.0.

5.0 Knowledge Engineering in WatQUAS 2.0

Many of the principles of knowledge engineering discussed in chapter 2 have been applied to the construction of the knowledge block of WatQUAS 2.0. This chapter outlines the methods with which the Expert System stores knowledge and the techniques that are utilized in WatQUAS 2.0 to extract knowledge from a variety of sources.

5.1 Incorporating Expert Knowledge into WatQUAS 2.0

The DBMS has been used whenever possible to store the domain and expert knowledge required by WatQUAS 2.0. This is to facilitate the access and modification of the information by the operators. The expansion of the knowledge base to contain comprehensive and extensive knowledge pertaining to a wide range of contaminants was one of the main goals in the development of the second version. The majority of the expert knowledge added to WatQUAS 2.0 is parameter specific.

The DBMS is external to the Expert System and the operation of WatQUAS is not necessary for editing of the parameter specific information. The DBMS is a "stand alone" system, independent of the Expert System.

A standard format for all contaminants is used to contain

the knowledge in the DBMS for all pollutants. The DBMS is divided into records, each record in the DBMS contains specific information relating to one category of knowledge for all pollutants. When the Expert System requires parameter specific information, it can locate the necessary knowledge by utilizing the parameter name and the name of the category of information required. For example, if the PWQO for aquatic life for the parameter lead is required, the Expert System retrieves this information from the DBMS. The same method is utilized to determine pollutant impacts, general information, toxicity ratings, etc.. The exact nature of the parameter specific information is described in the next section.

5.2 Parameter Specific Expert Knowledge

There are many sources for contaminant specific information, the chemical and general knowledge for most pollutants is relatively accessible. The major difficulty is finding information regarding pollutants in the aquatic environment. The Canadian Water Quality Guidelines (CWQG's) by the [Canadian Council of Resource and Environment Ministers 1987] is a thorough review of the nature of contaminants in the aquatic environment. Information pertaining to guidelines for human consumption, aquatic and plant life, and agricultural and industrial use

are presented. There is also information regarding contaminant;

* Uses and Production,
* Sources and Pathways for Entering the Aquatic Environment,
* Environmental Concentrations,
* Forms and Fate in the Aquatic Environment.

The pertinent information from the CWQG's for pollutants of concern in Ontario has been incorporated into the knowledge base of WatQUAS 2.0.

Since WatQUAS is for exclusive use in Ontario, the Provincial Water Quality Objectives for Ontario take precedence over the CWQG's. The CWQG's are more comprehensive than the Ontario PWQO's for Ontario, and are included in the knowledge base if the Canadian guideline is more stringent than the Ontario standard.

Figure 5.1 contains the parameter specific information for a typical contaminant for WatQUAS 2.0. The first category, "symbol", refers to the MOE laboratory designation of the pollutant. Information that identifies the contaminant is contained under "full name" and "abbreviation". Many contaminants are identified by more than one name, all common names are included if this is the case. The

***symbol=**Pb
***full name=**lead
***Abreviation=**PBUT mg/l
PBUR ug/filter
PBFT mg/l
***group=**con
***seriousness=**t
***MDC(overall)=**.01 mg/l
***MAC(drinking water)=**.05 mg/l
***RMPV=**default
***MAC(recreation)=**default
***MAC(Aquatic life)=**.01 mg/l
***MAC(Industrial)=**default
***Classification=**metal
***Likely Sources=**Industrial
=Urban
=Mining
=Natural
=Municipal
***Chemical description=**Toxicity dependant on alkalinity, increases as alkalinity increases. Chemical speciation of lead compounds is complex. In the aquatic environment, lead may be complexed with orgnic ligands, yielding soluble, colloidal, and particulate compounds. Sulphides,sulphates,oxides,carbonates and hydroxides of lead are insoluble.
***Fate=**Soluble lead is removed through association with sediments and suspended particultes, such as organic matter, hydrous oxides and clays. Sorption is the dominant mechanism controlling the distribution of lead in the aquatic environment. Lead is bioaccumulated by aquatic organisms,plants,inveterbrates and fish.
***Rec Drinking Water Treatment=**Conventional coagulation or lime softening is effective. Alum coagulation was found to achieve removals of 60-80% at low pH (6.5-&) and >90% at high pH (>9.5).
***Rec sampling Technique=**sediment sampling recommended
***Human Health Impact=**H
***Aesthetic Impact=**default
***Aquatic Impact=**H
***Socio Economic Impact=**H
***MISA Class=**p1
***CMR**
***atoral=**2
***atdermal=**
***ataquat=**7
***carcin=**7
***mutat=**
***terat=**3
***perwat=**
***persed=**
***bioaccbcf=**

Figure 5.1 Parameter Impact Knowledge Contained in the Knowledge Base

"seriousness" heading contains a toxic or non-toxic (to humans) designation for each parameter.

A series of guidelines for the pollutant is used to quantify the seriousness of the contaminant. "MDC" refers to the maximum desirable concentration, this is the strictest guideline that was found. Values for the following guidelines for the maximum acceptable concentration (MAC) are included if the information was available;

* "MAC (drinking water)",
* "MAC (recreational)",
* "MAC (aquatic life)",
* "MAC (industrial)".

The seriousness of a contaminant is also reflected by the category, "RMPV" (recommended maximum percentage of violations). If the water usage at the site includes drinking water then the pollution problem at the site is assessed in part by utilizing the RMPV. Extremely hazardous contaminants are assigned a low number for the recommended percentage of violations of the drinking water standard. Less hazardous pollutants are permitted higher percentages of violations. A zero tolerance contaminant would have a RMPV of 0.

The classification of the parameter identifies the class of pollutants it belongs to. Common classifications are conventional, organic, inorganic, radioactive and bacteriological. The group category identifies the group a contaminant belongs to inside a classification. For example, the groups nutrient, heavy metal, and trace metal are all contained in the inorganic classification.

The information from "likely sources" assists WatQUAS in identifying the origins of a contaminant. The major uses of the chemical are listed, some common uses are agricultural, industrial or municipal. The "fate" category permits the Expert System to assess the future of a pollutant in a stream. Contaminants that decay or dissipate relatively quickly are of less concern than pollutants which accumulate in the environment. Bioaccumulation and biomagnification knowledge is also contained in this category.

WatQUAS utilizes the information from "Recommended Drinking Water Treatment" to determine the necessary treatment to remove a pollutant from drinking water. The Expert System recommends a simple abatement strategy for a pollutant by analyzing source information and the pollutant grouping. For example, if bacteriological pollution was a problem and there was an STP upstream then the measures recommended

by WatQUAS 2.0 would be to investigate the STP and to possibly chlorinate the STP effluent.

The "Recommended Sampling Technique" category supplies WatQUAS 2.0 with information pertaining to the most suitable methods to monitor a contaminant. If a parameter accumulates in sediment then the knowledge base informs WatQUAS that sediment sampling is recommended. If the effects of a contaminant are unknown or not fully understood then acute toxicity testing may be recommended.

The next group of categories contain information regarding the different impacts of the pollutant. The various impacts are rated as high, moderate, low, or default (for unknown) depending upon the individual parameter. The four impact classifications are;

* Human Health Impact,
* Aesthetic Impact,
* Aquatic Impact,
* Socio-Economic Impact.

The remaining information is hazard and toxicity data from the EMPPL of the Ontario MOE. Section 5.3 outlines the form of this knowledge and the methods with which WatQUAS 2.0 makes use of it.

5.3 Hazardous Contaminant Assessment

WatQUAS 2.0 must be capable of identifying the hazards associated with contaminants in the environment. The knowledge block of the Expert System must contain comprehensive and exact information concerning the dangers each pollutant represents to human and aquatic life. There are many thousands of different chemicals and substances that are potential water quality pollutants. The construction of a knowledge base containing detailed knowledge of the hazards from all possible parameters is a prohibitive task. WatQUAS 2.0 will rarely be required to utilize the hazard information for many of the contaminants in the knowledge base. However, when information regarding the hazards of rare pollutants is required, the Expert System will have access to it.

5.3.1 Effluent Monitoring Priority Pollutants List

In 1987, the Ontario Ministry of the Environment produced a list of hazardous contaminants, from the municipal and industrial sector that could potentially be found in Ontario waterways [MOE EMPPL 1987]. This list of pollutants, the Effluent Monitoring Priority Pollutants List (EMPPL), was published in conjunction with the MOE; Municipal and Industrial Strategy for Abatement policy (MISA).

For a contaminant to be assigned to the EMPPL, it must be deemed hazardous and have been discovered or be potentially present in industrial or municipal discharge effluents in Ontario. The EMPPL is the basis for developing regulations for the industrial and municipal sector with regards to specific effluent discharge to the environment . The EMPPL is mostly composed of organic chemicals and toxic metals.

The EMPPL is subdivided into two groups; a primary list and a secondary list. The primary list is composed of chemicals that have been detected in the Great Lakes or in industrial or municipal waste effluent. The secondary list contains pollutants that are considered hazardous and which may be present in effluents, but have not been detected in effluents originating in Ontario or in the environment of Ontario. Pesticides and many conventional pollutants have not been included in this version of the EMPPL. Subsequent revisions to the list are expected to increase the number of pollutants monitored by the program.

Since the purpose of WatQUAS is to assess water quality in Ontario rivers, all pollutants listed in the MOE EMPPL are contained in the knowledge block of WatQUAS. The EMPPL also supplies comprehensive toxicity and hazard information for each parameter. This knowledge is in the form of numeric scores for each category of hazard. The knowledge base of

WatQUAS 2.0 contains the scores for all the parameters on the EMPPL. A separate module in the WatQUAS 2.0 Expert System interprets what the numerical scores represent.

Pesticides, conventional, bacteriological and radioactive pollutants and many non-toxic pollutants not on the EMPPL are also included in the knowledge block of WatQUAS. This results in a knowledge base which contains the necessary information to assess any pollutant possibly found in Ontario rivers. However, contaminant hazard ratings are not available for these pollutants

The first edition of the EMPPL contained 180 chemicals which were judged to be environmentally hazardous. The possible sources of the chemical and the hazard ratings of the pollutant are also contained in the list. The source and toxicity information for each parameter on the EMPPL was assembled by the MOE using data from a variety of sources. The Michigan Department of Natural Resources; Critical Materials Register, 1980 (CMR) was used because of the proximitry of Michigan to Ontario and the similar type of industrial development in the two areas. The Niagara River Toxics Committee (NRTC) reviewed and investigated pollutants specifically identified in the Niagara River. The remainder of chemicals on the list were reviewed and investigated under the supervision of the MOE. The three

agencies each utilized a different criteria for rating the toxicity of pollutants.

5.3.2 CMR for Michigan

The MISA program of Ontario utilized the 1980 version of the CMR of Michigan to assess many contaminants. A total of 223 compounds were contained in this list [MOE 1987]. Only the compounds potentially found in Ontario were included in the EMPPL for the MISA program. The chemicals reviewed in the CMR were rated on the basis of;

* Persistence,
* Bioaccumulation,
* Acute Toxicity,
* Hereditary Mutagenicity,
* Teratogenicity,
* Carcinogenicity,
* and Other Adverse Effects.

The pollutants were rated on a scale of 0 - 7 (0 best - 7 worst) for each area of concern, except for persistence which was rated from 0 - 4. Figures 5.2 a & b contain the exact breakdown of the necessary ratings that a substance must receive before being promoted to the CMR primary or secondary list.

CMR Criterion	Concern Level
Persistance	≥ 1
Bioaccumulation	≥ 3
Acute Toxicity	≥ 3
Other Adverse Effects	≥ 3
Hereditary Mutagenicity	≥ 4
Teratogenicity	≥ 3
Carcinogenicity	≥ 2

Figure 5.2a

Criteria for Promotion from

Primary Group to EMPPL

CMR Criterion	Concern Level
Persistance	≥ 4
Bioaccumulation	≥ 7
Acute Toxicity	≥ 7
Other Adverse Effects	≥ 7
Hereditary Mutagenicity	≥ 7
Teratogenicity	≥ 7
Carcinogenicity	≥ 7

Figure 5.2b

Criteria for Promotion from

Secondary Group to EMPPL

5.3.2.1 Acute Toxicity

Acute toxicity assessment is sub-divided into three types of toxicity; oral, dermal and aquatic. Oral acute toxicity assesses the dosage of a substance that is toxic through ingestion. Dermal acute toxicity assesses the toxic dosage of a substance which is contracted through skin contact. Aquatic acute toxicity assesses the quantity of a substance detrimental to aquatic life. Scores for each type of acute toxicity are based upon the levels of lethal dosages or lethal concentrations.

An overall acute toxicity rating is assigned from the highest score from any of the three types of toxicities. For example, if oral was assigned a score of 7 while dermal and aquatic were assigned a score of 2 each, the overall acute toxicity rating would be 7 (based on the oral score).

5.3.2.2 Carcinogenicity

Identification and control of carcinogenic chemicals in the environment is necessary in order to control the incidence of cancer. The following rating scale is used to define the carcinogenicity of a substance [MOE 1987];

SCORE	CATEGORY
7	The chemical has been demonstrated to be human positive, potential human or animal positive carcinogen by the oral or dermal route of exposure.

3 The chemical has been demonstrated to be a potential animal carcinogen by the oral or dermal route of exposure.

2 The chemical has been demonstrated to be an animal positive or potential animal carcinogen by any route other than oral or dermal; or has been demonstrated by accepted mutagenicity screening tests or accepted cell transformation studies to be strongly suspect carcinogen.

1 The chemical has been demonstrated by accepted mutagenicity tests or accepted cell transformation studies to be a suspect carcinogen.

0 The chemical has been tested by the above systems and has not been demonstrated to cause cancer or to be a suspect carcinogen.

5.3.2.3 Hereditary Mutagenicity

Hereditary mutagenicity is an effect discernible only over long periods of time. Many generations of a species are often required to be tested in order to discover any mutagenic effects caused by a substance. The following rating scale is used to score hereditary mutagens [MOE 1987};

SCORE	CATEGORY
7	Confirmed hereditary mutagen
4	Potential hereditary mutagen in multicellular organisms
2	Potential hereditary mutagen in micro-organisms
0	Not demonstrated to be a hereditary mutagen

The CMR defines a hereditary mutagen as a chemical which

produces a statistically significant dose related mutagenic effect in test micro-organisms without the use of metabolic activators or in subsequent generations of the micro-organism. In complex multicellular animals, hereditary mutagens are substances which produce mutations inheritable in subsequent generations of the test organism.

5.3.2.4 Teratogenicity

The CMR defines a teratogen as a substance which causes alterations in the formation of cells, tissues, and organs resulting from physiologic and biochemical changes. The following rating system is used to score teratogens [MOE 1987];

SCORE	CATEGORIES
7	Confirmed Teratogen
3	Potential Teratogen
0	Not Teratogenic

To be classified as a teratogen a chemical must be confirmed or be potentially shown to be teratogenic in one animal species by oral or dermal exposure routes.

5.3.2.5 Persistence

The Michigan CMR considers persistence in the environment to be an important property because of the long term

effects any continual exposure to a substance could have on organisms. Long term persistence also indicates that there is a greater risk of contact or exposure with the chemical. The following scoring system based on the estimated half life in soil or water of the chemical is utilized to rate the substance [MOE 1987];

SCORE	CATEGORY	HALF LIFE (weeks)
4	Very Persistent	> 52
3	Persistent	40 - 52
2	Slowly Degradable	27 - 39
1	Moderately Degradable	14 - 26
0	Readily Degradable	0 - 13

5.3.2.6 Bioaccumulation

The CMR uses the partition coefficients for n-octanol/water as a measure of the tendency for an organic compound to transfer from water to organisms and bioaccumulate. The n-octanol/water partition coefficient, P, is defined as the ratio of the concentration of a compound in octanol to its concentration in water. P is usually expressed as the base 10 log of the partition coefficient. The following rating system and partition coefficients are used to score a pollutant for bioaccumulation [MOE 1987];

SCORE	BIOACCUMULATION	LOG P
7	>= 4000	>= 6.00
3	1000 - 3999	5.00 - 5.99
2	700 - 999	4.50 - 4.99
1	300 - 699	4.00 - 4.49
0	< 300	< 4.00

5.3.2.7 Other Adverse Effects

This category is divided into three subsections; terrestrial animals, aquatic organisms and plants. The following scoring system is utilized for rating the seriousness of the effects of a substance on terrestrial animals [MOE 1987];

SCORE	CATEGORY
7	Produces an irreversible effect at a very low dose by oral or dermal routes.
3	Irreversible effects during or following cessation of the low level exposure by oral or dermal routes.
2	Reversible effects following cessation of low level exposure by oral or dermal routes.
1	Adverse effects by inhalation route.
0	No detectable adverse effects.

Other adverse effects for terrestrial animals covers a wide range of effects following contact with a substance. Some of these effects are; Benign neoplasis, embryo or fetal

mortality, metabolic disorders, cataracts, cirrhosis, sterility, vitamin deficiencies, skin or eye irritation to name a few.

Adverse effects on aquatic organisms include stresses on the reproductive cycle and other sub-lethal problems. The following scores rate the effects on aquatic organisms of a substance [MOE 1987];

SCORE	MEDIAN EFFECTIVE CONCENTRATION (EC-50)
7	< 0.1 mg/l
3	> 0.1 - 1 mg/l
2	> 1 - 10 mg/l
1	> 10 - 100 mg/l
0	> 100 mg/l

Adverse effects on plant life is a concern because of the potential for contaminated water to be used for irrigation purposes and the need for plant life to maintain a healthy environment. Plant effects for a contaminant are scored in the following manner [MOE 1987];

SCORE	WATER
3	< 0.5 mg/l
2	> 0.5 - 5.0 mg/l
1	> 5 - 50 mg/l
0	> 50 mg/l

5.3.3 Niagara River Toxics Committee Assessment Criteria (NRTC)

The Niagara River Toxics Committee assessed contaminants found in the Niagara River, the eastern end of Lake Erie and the west end of Lake Ontario. The committee assessed 267 various chemicals, most of which have been identified in other parts of the province. The NRTC utilized a ranking system based on information from the International Joint Commission Health Effects Committee (HEC) report and the Acute Effects Ranking (AER) system (an adaptation of the Michigan CMR)

All contaminants were divided into one of three major groups. Group I pollutants are the most serious and require immediate attention. There are seven subsections of group II, group IIA are substances with a slightly lower priority than group I pollutants. The other group II subsections, B - G, are ranked in decreasing order of priority. Group III substances have very low priority and do not require immediate attention.

The NRTC criteria for rating the hazards of contaminants is very general and not as detailed as the CMR or EMPPL criteria. Only seven contaminants on the EMPPL were assessed by the NRTC.

5.3.4 EMPPL for Ontario

The Ontario MOE utilized a similar scoring system as the Michigan CMR for rating hazardous substances which had not previously been assessed by the CMR or NRTC. The Ontario scoring system utilizes scores ranging from 0 to 10 (0 best - 10 worst) for all categories except environmental transport which is assigned a maximum value of only 4. The categories defined as areas of concern from the presence of pollutants in the environment are;

* Environmental Transport,
* Environmental Persistence,
* Bioaccumulation,
* Acute Lethality,
* Sub-Lethal Effects on Non-mammalian Animals,
* Sub-Lethal Effects on Plants,
* Sub-Lethal Effects on Mammals,
* Teratogenicity,
* Genotoxicity/Mutagenicity,
* Carcinogenicity.

These categories have previously been defined in section 5.3.2 with the outline of the CMR of Michigan. Figure 5.3 shows the ratings for each category at which a parameter is considered hazardous and thus included in the EMPPL.

MOE Criterion	Concern Level
Persistance	≥ 7
Bioaccumulation	≥ 7
Acute Lethality	≥ 6
Sub-Lethal Toxicity Non-Mammalian	≥ 6
Sub-Lethal Toxicity Plant	≥ 6
Sub-Lethal Toxicity Mammalian	≥ 6
Mutagenicity/Genotoxicity	≥ 6
Teratogenicity	≥ 0
Carcinogenicity	≥ 2

Figure 5.3a Criteria for Promotion
from Primary Group to EMPPL

MOE Criterion	Concern Level
Persistance	≥ 10
Bioaccumulation	≥ 7
Acute Lethality	≥ 8
Sub-Lethal Toxicity Non-Mammalian	≥ 6
Sub-Lethal Toxicity Plant	≥ 10
Sub-Lethal Toxicity Mammalian	≥ 10
Mutagenicity/Genotoxicity	≥ 10
Teratogenicity	≥ 4
Carcinogenicity	≥ 6

Figure 5.3b Criteria for Promotion
from Secondary Group to EMPPL

WatQUAS 2.0 possesses knowledge concerning the known hazards of many of the pollutants potentially found in Ontario. An expert assessment of the problems that contaminants present in a stream is completed using the EMPPL hazard ratings.

5.4 Heuristics in WatQUAS 2.0

Many of the rules in WatQUAS 2.0 are simple frames and are utilized for all parameters. The similar format of the rules for each parameter was recognized in the construction of version two and rule frames were developed. For most water quality situations only one general rule frame with the specific information being retrieved from the DBMS was used for all parameters. Figure 5.4 illustrates a series of rules for various parameters from WatQUAS 1.0 that pertain to the same water quality assessment area. Figure 5.5 shows the form of the same rules in WatQUAS 2.0. The variables (words with an "&" prefix) in the rule of figure 5.5 are required to retrieve the correct information from the DBMS.

The expansion of the rules of the Expert System is an area which requires substantially more work to encompass more areas of water quality assessment.

-- RSP rules

```
rule RSP_setup
(goal function=assess; object=water_quality;
status=active);
&1(parameter abbreviation=RSP; class=||)
-->
modify &1(class = solid;
  human_health-impact = low;
  aesthetic_impact = moderate;
  aquatic_impact = moderate;
  socio_economic_impact = high;
  dissipation = seasonal;
);
```

-- NN03FR rules

```
rule NN03FR_setup
(goal function=assess; object=water_quality;
status=active);
&1(parameter abbreviation=NN03FR; class=||)
-->
modify &1(class = nutrient;
  human_health_impact = moderate;
  aesthetic_impact = moderate;
  aquatic_impact = moderate;
  socio_economic_impact = moderate;
  dissipation = seasonal;
);
```

-- turbidity rules (TURB)

```
rule TURB_setup
(goal function=assess; object=water_quality;
status=active);
&1(parameter abbreviation=TURB; class=||)
-->
modify &1(class = physical;
  human_health_impact = low;
  aesthetic_impact = moderate;
  aquatic_impact = moderate;
```

Figure 5.4 Rules from WatQUAS 1.0

```
--   WatQUAS version 2.0
--   (For VAX and MicroVax II Unix/Ultrix)
--
--   (C) Copyrighr 1987
--   W.C. Allison
--   University of Waterloo
--   Waterloo, Ontario
--
--   File:  parsum_rules.ops
--
--   Function: parameter-summary rules
--

module parsum_rules ()
{

use definitions;

-- parameter summary

rule PARAMETER_SUM
{

        (goal function=assess; object=parameter_sum;
        status &1 (parameter=&parabb);
        -->
        write ()  :Parameter Summary for:, &1.parameter,
        write ()  :Full Name is:, &parname, '/n';
        write ()  :Unit of Measurement is:, uofm, '/n';
        write ()  :class = :,&class, '/n';
        write ()  :dissipation = :,&diss, '/n';
        write ()  :human_health_impact = :,&hhi,'/n';
        write ()  :aesthetic_impact = :,&aesi, '/n';
        write ()  :aquatic_impact = :,&aquati, '/n';
        write ()  :socio_economic_impact = :,&sei, '/n';
        write ()  '/n';
        write ()  :Chemical description:  :,&chemdesc, '/n/;
        write ()  '/n';
        };
```

Figure 5.5 WatQUAS 2.0 Rule Frame

6.0 Recommendations and Future Work

Version 2.0 of WatQUAS remains a small, skeletal expert system which requires extensive work to complete. The software package must be completed and graphics routines developed. This work requires the expertise of a highly skilled computer scientist. Although WatQUAS 2.0 contains more knowledge and rules and is more versatile than the prototype version, it will still incapable of handling many situations commonly encountered in water quality assessment. A plan for future work and recommendations of water quality assessment areas that could be developed for the WatQUAS Expert System are contained in this chapter.

6.1 Knowledge Expansion and Enhancement

The most important feature of any Expert System is its knowledge block and heuristics. Theoretically, an IKBS is supposed to contain all the information that an expert could require to assess a given situation. Experts acquire large quantities of knowledge throughout their life time. Many years of experience and training are usually required prior to a person achieving an "expert" status. Similarly, the knowledge base for an Expert System requires many years of development and numerous revisions before it can be considered an "expert" in its field.

A human acquires new information and expands his/her knowledge base continually. An expert system must be similarly expanded in order that the user is assured of receiving the best and most "up to date" response possible from the system. WatQUAS 2.0 utilizes a DBMS in which large quantities of parameter and situation specific knowledge may be readily accessed by the Expert System. MOE personnel can continually expand the knowledge base of WatQUAS through the DBMS facility.

6.1.1 Site Specific Knowledge

There is a lack of recorded and catalogued site specific knowledge regarding the area surrounding water quality monitoring stations in Ontario. Most information concerning the monitoring stations and the surrounding area is possessed by individual MOE contract samplers, local MOE personnel and the conservation authorities.

The site specific knowledge concerns such areas as;

* Site geography,
* Background contaminant levels,
* Site sensitivity to various contaminants,
* A detailed profile of local and up-stream polluters,
* Local and down-stream water usage.

The quality of the water quality assessment of a river,

produced by WatQUAS, could be improved greatly if comprehensive knowledge of local conditions were available to the expert system.

6.1.2 Pollutant Specific Knowledge

WatQUAS 2.0 recognizes all of the contaminants on the MOE EMPPL and many conventional and bacteriological pollutants. Although the Expert System may recognize the pollutants, there is still many gaps in the knowledge base. Future work on WatQUAS could entail completing the information profiles of the water quality pollutants.

There are still many contaminants that the Expert System does not recognize. There is no knowledge concerning many pollutants from such categories as;

* Pesticides,
* Herbicides,
* Radioactive pollutants,
* Hazardous organic contaminants.

It is a prohibitive task to incorporate knowledge concerning all of the water pollutants from these categories into WatQUAS. Over a period of a number of years many of the contaminants could be added to the Expert System.

6.1.3 Pollutant Interaction Knowledge

Pollutant interaction knowledge is a form of pollutant specific knowledge. It concerns the overall effects and chemistry of the combination of two or more pollutants in the aquatic environment. This interaction of pollutants is referred to as synergy. There is very little technical information available concerning the synergy of chemicals in the aquatic environment. A great deal of the knowledge regarding the interaction of water quality pollutants can be derived from water quality and chemistry experts.

It would be a prohibitive task to catalogue knowledge pertaining to all of the potential combinations of two or more water quality pollutants. A far more reasonable goal is to determine the synergy of the most common environmental contaminants.

6.1.4 Problem Specific Knowledge

An Expert System for water quality assessment must be capable of recognizing specific pollution problems. WatQUAS should contain expert knowledge which will permit it to determine the water quality problem by analyzing the effects of the pollutants in the stream. A simple example is; WatQUAS should recognize that prolific plant growth is an indicator of nutrient pollution. There are many

instances where the ambient stream conditions point to a specific pollution problem. More of this type of knowledge should be derived from water quality experts and incorporated into the Expert System.

Future development of WatQUAS can focus on programming the Expert System to teach itself problem specific knowledge. WatQUAS can utilize the expert assessment of a water quality situation it has completed to assist itself in a subsequent analysis of a similar situation. In the artificial intelligence field this is termed "learning". An Expert System stores the results and interpretations of a given situation and retrieves them for reference when a similar situation arises in the future.

6.1.5 WatQUAS as a General Information Provider

An extra benefit of WatQUAS 2.0 possessing a large knowledge base is that it can permit non-experts access to expert information. The knowledge block in WatQUAS 2.0 has been developed with the emphasis on information pertaining to water quality. Chemical and technical manuals often contain much extraneous information that a hydrologist or water quality technician would not find useful. WatQUAS presents information that is relevant to a person examining water quality.

Future work on the Expert System could encompass the installation of a facility to permit the easy access of water quality knowledge. This facility would require a search technique in the DBMS and a natural language processor to interface with the user.

6.1.6 Help Facilities

Extensive testing and operation of WatQUAS will indicate areas in which the user will encounter problems and difficulties. The HELP facility of the Expert System should be expanded to assist the operator with any problems that could be encountered.

The HELP facility will also benefit from the inclusion of a natural language processor software package. This will enable the operator to communicate efficiently with the Expert System.

6.2 River and Basin Assessment

Future versions of WatQUAS should be constructed such that the water quality assessment of an entire river or basin is possible. The DBMS permits the Expert System access to the historical time series record for any site. The expert assessment of the water quality analysis of a site is also stored in the DBMS. Modules which can interpret the results of analyses of the water quality by WatQUAS at

related sites should be developed. This strategy for river assessment will only be applicable if the water quality time series record utilized by WatQUAS contains data for more than one site on the river.

By assessing an entire river, the Expert System can determine problem areas in the stream and can identify sources of pollution. Knowledge can be incorporated into WatQUAS which will permit recommendations for effective abatement strategies and control options given the pollution problem for the entire river.

The assessment of the water quality monitoring sites throughout an entire basin will permit the Expert System to identify pollution "hot spots" in the basin. "Hot spots" are localized areas of high pollution levels. The type of knowledge WatQUAS should contain to deal with problem areas is;

* Recommending additional sampling programs,
* Identifying possible sources,
* Recommending control options,
* Assessing the potential of pollution spread and migration,
* Determining the long term effects and impacts of the pollution in the basin.

The DBMS provides a basis enabling WatQUAS to assess more

than one site. Most of the work in this area must be focused on developing rules and incorporating expert knowledge into river and basin assessment modules.

6.3 Graphics

The extent to which graphics may be utilized in the Expert System will depend upon the type of computer operating the WatQUAS system. Ideally, a computer with full graphic capabilities and a high resolution colour monitor should be utilized. Personnel from the Hydrology Unit of the MOE have indicated that displaying the results of the entire river and basin assessment and identifying problem areas would be very beneficial.

Geographical maps of Ontario rivers and basins which locate water quality monitoring stations can be programmed into WatQUAS. The maps can display pollutant levels, trends, or violations for each site located on the river or basin. The severity of pollution problems at each site can be colour coded to permit the operator to instantly grasp the pollution situation over a large area.

Other areas where additional graphics could be beneficial to the WatQUAS Expert System are;

* Displaying enhanced water quality regression techniques,

* Plots showing the expected effectiveness of control options,

* Plots displaying the results of modeling (section 6.4).

Graphics is a feature that is not integral to the Expert System, most of the results produced by WatQUAS may be displayed graphically. The developer has to only specify the proper external plotting routines to display the desired results.

6.4 Statistical and Simulation Models

Many types of water quality assessment techniques require continuous time series quality data. This is rarely available for sites in the water quality monitoring network of the MOE. A continuous quality record can be constructed from existing discrete samples by utilizing statistical or simulation models.

Expert Systems have been specifically designed to calibrate and validate different types of models. Future versions of WatQUAS should include modules which can conduct an expert calibration and validation of simple statistical or simulation models using the quality record from the MOE monitoring sites. Expert techniques to calibrate models which simulate the hydrology and quality of rivers or basins could also be developed.

WatQUAS can utilize calibrated models to determine the effectiveness of abatement strategies and control options before they are implemented. Determining pollution trends and identifying potential problems can also be accomplished by utilizing a calibrated model.

6.5 Purpose Dependent Expert Systems

Individual versions of WatQUAS should be developed for specific users. The Expert System can be tailored to suite the purposes and needs of a particular group of operators. A small, simplified version of WatQUAS could be developed for users who only have access to a basic computer system and will be concentrating mostly on water quality assessment. This edition of WatQUAS would possess;

* The water quality record only for the sites requiring assessment,
* Knowledge concerning only the contaminants potentially found in the area,
* Knowledge concerning only the sites in the area,
* limited graphics capability.

This basic Expert System would be ideal for placement in MOE regional offices and with the local conservation authorities. WatQUAS could assist with the expert interpretation of water quality data and provide users with expert knowledge pertaining to the localized area of stream

assessment.

The full version of WatQUAS could be utilized by hydrologists responsible for assessing river quality province wide and concerned with a wide range of water quality problems. A fully configured micro-computer would be required to operate the comprehensive Expert System. The micro-computer system should consist of;

* A 386 co-processor,
* 2 - 5 Megabytes of RAM,
* 70 - 110 Megabyte hard-disk,
* High Resolution Colour Monitor,
* One 5.25 inch and One 3.5 inch disk drive.

The IBM Personnel System 2, Model 80 with the necessary accessories would be an ideal choice to operate the full WatQUAS Expert System. The subsequent development and expansion of WatQUAS could be accomplished on this system. The basic Expert Systems for distribution could be configured with the principal computer system to operate on the smaller systems.

Ideally, all computers utilizing the WatQUAS Expert System should be linked together. By linking the computers, all users could benefit from subsequent expert knowledge being added to one version of WatQUAS. Linking could be

accomplished easily and cost effectively by utilizing BELL telephone lines and modems.

6.6 Testing and Evaluation

WatQUAS 2.0 must be subjected to intensive testing and evaluation. The majority of this testing will occur after the Expert System has been transferred to the MOE and in conjunction with personnel from the Hydrology Unit. The results from this procedure will indicate weaknesses in WatQUAS 2.0 and areas that require further refinement. The evaluation of the expert knowledge of WatQUAS 2.0 will indicate additional specific knowledge that is required.

7.0 Summary and Conclusions

This section summarizes the work completed to date by the author on the WatQUAS Expert System. Conclusions pertaining to the second version of WatQUAS are also presented.

7.1 The Knowledge Base

Originally, WatQUAS 1.0 contained knowledge concerning twelve pollutants. The knowledge addressed approximately ten areas of concern and was of a general nature. As a result of this work, WatQUAS 2.0 now possesses knowledge relating to 255 various water quality pollutants. The knowledge addresses approximately 50 areas of concern. Due to the unavailability of specific information, all 50 areas of knowledge are not complete for every parameter. A comprehensive and thorough data base has been completed. The Canadian Water Quality Guidelines document was one source of the information relating to contaminants in the aquatic environment.

The current knowledge base has now been developed through the use of a Data Base Management System (DBASE III). It is an organized and easily accessible knowledge base, that permits rapid modification by the user. Subsequent updating and expansion of the knowledge base will require

minimal computer expertise.

7.2 Water Quality Assessment Techniques

The water quality assessment techniques utilized by WatQUAS 1.0 have been expanded and enhanced as a result of this work. A non-parametric statistical analysis module was constructed for inclusion in WatQUAS 2.0. This permits water quality data, regardless of its nature or distribution, to be accurately and thoroughly analyzed. A new statistical module also inspects for and manages outliers in the water quality data distribution. WatQUAS 2.0 is assured of utilizing a valid time series record.

The violation assessment techniques of WatQUAS 1.0 have been changed for the second version. Originally, if a PWQO was unspecified, then the 90th percentile of the time series record was utilized as the value for the water quality objective. WatQUAS 2.0 considers the seriousness of the pollutant, before arbitrarily assigning a value to the objective. If the pollutant is toxic and a PWQO is not specified then any detection of the contaminant is considered a violation by WatQUAS. This technique always yields a conservative violation assessment for a hazardous contaminant.

A Cumulative Distribution Function module was incorporated

into WatQUAS 2.0. This permits the probability that a water quality sample will be in violation of the stream standard to be determined. Parameteric and non-parametric CDF techniques are utilized by the Expert System.

7.3 Water Quality Indices

WatQUAS 2.0 employs two new water quality indices. A "Prevalence, Duration and Intensity Index" is utilized to account for conditions at the the water quality monitoring site. The second index examines and aggregates individual pollutants based on the seriousness and impact on the environment that each contaminant represents. Although, water quality indices are not recognized as a completely reliable tool for the measurement of water quality, they are well suited for computer application in an Expert System. The indices permit WatQUAS 2.0 to examine many water quality situations which otherwise would require a prohibitive quantity of expert and domain knowledge.

7.4 Pollutant Loadings

WatQUAS 2.0 utilizes the BEALE ratio estimator for the calculation of pollutant loads. This is the same method utilized by the Ontario Ministry of the Environment. The ratio estimator permits the Expert System

to calculate loads which are more accurate than those calculated by WatQUAS 1.0 and are fully compatible with the MOE.

The ratio estimator technique also permits WatQUAS 2.0 to identify pollution sources. The quantity of point-source pollution is calculated in the flow stratum representing base flow. Once identified, the quantity of point-source pollution is subtracted from the total quantity of pollution in non-base flow strata to yield the total non-point source pollution load.

Hypothetical pollutant load reduction is also examined by WatQUAS 2.0. Revised pollutant load estimates are calculated by utilizing a percentage reduction in pollution, supplied by the user to the Expert System. WatQUAS 2.0 can examine point source and non-point source pollution reductions. This permits water quality managers to estimate the effectiveness of various pollution control and abatement strategies.

7.5 Hazardous Contaminant Assessment

WatQUAS 2.0 can accurately assess the hazards presented by approximately 255 various water quality contaminants. The hazard assessment information for pollutants encountered in Ontario was derived from the MISA EMPPL study. The Expert

system interprets the ratings from the EMPPL study in order to achieve a comprehensive assessment of the hazards an individual contaminant represents.

7.6 Expert Assessment

WatQUAS 2.0 utilizes rule frames for the expert assessment of a water quality situation. In conjunction with the DBMS managed knowledge base, only one rule frame is required for all 255 contaminants for each water quality situation. Rule Frames eliminated the problem encountered by WatQUAS 1.0 of having to write an individual rule for each of the monitored parameters.

The second version of the Expert System contains modules which have the capability to investigate such situations as;

* Drinking Water Assessment
* Recreation Usage Assessment
* A Pollutant Summary
* Seasonality of Data Assessment
* Parameter Hazard Assessment
* Water Quality Index Assessment
* Violation Assessment
* Trend Assessment
* Control Measure Assessment
* Fate Assessment
* Source Investigation

7.7 WatQUAS Operation

WatQUAS 2.0 is composed of menus to allow for the operation of the Expert System. The user is presented with menus

which permit him/her to access the various aspects of WatQUAS 2.0.

WatQUAS 2.0 is designed to operate on a IBM micro-computer. Upon completion of programming and debugging, it can be installed in MOE offices throughout the province.

7.8 Conclusion

In conclusion, WatQUAS 2.0 remains a relatively small expert system. Work by a computer programming specialist is required to complete this version. Many years of effort are still necessary to make it a truly comprehensive and encompassing tool for water quality assessment. Hopefully, this project will be continued and the WatQUAS Expert System expanded so that it may assist water quality managers throughout the province.

GLOSSARY

(AER)	Acute Effects Ranking
(CDF)	Cumulative Distribution Function
(CMR)	Critical Materials Register
(Cs)	Coefficient of Skew
(CWQG)	Canadian Water Quality Guidelines
(DBMS)	Data Base Management Systems
(EMPPL)	Effluent Monitoring Priority Pollutants List
(ETMP)	Enhanced Tributary Monitoring Program
(HEC)	Health Effects Committee
(IJC)	International Joint Commission
(IKBS)	Intelligent Knowledge Based System
(I/O)	Input/Output
(IQR)	Interquartile Range
(KBS)	Knowledge Based System
(M)	Stream Length
(MAC)	Maximum Acceptable Concentration
(MDC)	Maximum Desirable Concentration
(MISA)	Municipal and Industrial Strategy for Abatement
(MOE)	Ontario Ministry of the Environment
(NRTC)	Niagara River Toxics Committee
(P)	Prevalence
(PDF)	Probability Distribution Function

(PDI)	Prevalence, Duration and Intensity
(PWQO)	Provincial Water Quality Objective
(RMPV)	Recommended Maximum Percentage of Violations
(STP)	Sewage Treatment Plant
(t 1/2)	Half-Life
(WQI)	Water Quality Index
(7 LQ 20)	The minimum seven day consecutive low flow with a 20 year return period.

APPENDIX A

WATER QUALITY EXPERTS

Water Quality Experts

Dr. Lloyd Logan, Ph.D., P. Eng.

Co-ordinator, Hydrology and Networks Unit, Water Resources Branch, Ontario Ministry of the Environment.

Dr. Logan's Curriculum Vitae is reproduced on the following pages.

Mr. Brian Whitehead, M.A.Sc.

Water Quality Specialist, Hydrology and Networks Unit, Water Resources Branch, Ontario Ministry of the Environment.

Dr. Byron Bodo, Ph.D.

Water Quality Specialist, Hydrology and Networks Unit, Water Resources Branch, Ontario Ministry of the Environment.

Education

Degree:

Ph.D. Engineering (Water Resources)
University of Waterloo, 1979

M.Sc. Engineering (Hydrology)
University of Guelph, 1968

B.Sc. Engineering (Soil and Water)
Techncian, I.I.T., Haifa, Israel, 1966

Other Training:

University of Toronto

- Stochastic Processes, 1970
- Operation Research and Management, 1970
- Power Spectral Density Analysis, 1971
- Simulation and Management Modelling, 1972

University of Nebraska

- Simulation of Water Resources Systems, 1971

Case Western Reserve University

- Hierarchical Approach in Water Resources Planning Management, 1976

Management Training

- Self, Social and Business Development, 1971-73
- Power Play, 1972
- Communication Workshop, 1973
- Management Development, 1975
- Project Management, 1977
- Management: A Systematic Approach, 1979
- Effective Writing, 1980
- Media Relation, 1983
- Performance Management, 1985
- Excellence in Thinking and Writing, 1986

Professional Affiliation:

- The Association of the Professional Engineers of the Province or Ontario
- The Canadian Society of Professional Engineers
- The American Geophysical Union
- International Association of Hydrological Sciences

Other Skills:

- Computer Programming and Analysis
- Public Speaking, CTM

Languages

- English
- Hebrew

Committee Membership

- Co-ordinating Committee for Canada/Ontario Agreement for Water Quantity Surveys
- Ontario Water Management Research and Services Committee
- Environmental Monitoring and Modelling Committee
- Atrazine Study Technical Committee
- Sturgeon - Rice Lake Study Technical Committee
- Ontario Waste Management Committee

Publications

. Bulletin	2
. Seminar/Workshop	9
. Report	7
. Scientific Paper	24
. Invited Papers	7
. Thesis (Ph.D.)	1

APPENDIX B

WatQUAS 2.0 Computer Modules

Water Quality Assessment Modules

1 Water Quality Index;General Site Index
Parameter Specific Index
2 Cumulative Distribution Function;Parametric Analysis
Non-parametric Analysis
3 Non-parametric Statistical Package
4 Outlier Identifier
5 Beale Load Estimator
6 Pollution Source Identification
7 Pollution Reduction Calculator
8 Violation Assessment

Expert Assessment Modules

1 Hazard Assessment
2 Drinking Water Assessment
3 Recreational Usage Assessment
4 Parameter Summary
5 Fate Assessment
6 Source Identification
7 Trend Assessment
8 Seasonality of Data Assessment
9 Violation Assessment
10 Control Measure Assessment
11 Impact Assessment

WatQUAS Operation Modules

1 Modify	7 List
2 Shell	8 Show
3 Describe	9 Stats
4 Summary	10 Identify
5 Help	11 Graph
6 Quit	

The listings of these computer programs and any supporting programs from WatQUAS 1.0 are available upon request.

REFERENCES

[Allen 1986] Allen, W.P., "An Expert System for Water Quality Assessment". Thesis presented to the University of Waterloo in fulfillment of the thesis requirements for the degree of Master of Applied Science in Systems Design Engineering, 1986.

[Ball 1980] Ball, R.O. and Church, R.L.,"Water Quality Indexing and Scoring". J. Environ. Engineer. Div., ASCE 106, 757-771, 1980.

[Berel 1987] Berel, A.C., "EXPERT SYSTEMS Strategic Implications and Applications". Halsted Press, Toronto, 1987.

[Bodo 1988] Bodo, B.A., "Basic Statistical Methods for Water Quality Data Analysis". Unpublished Paper, Ontario Ministry of the Environment 1988.

[Bodo & Unny] Bodo, B., and Unny, T.E., "Sampling Strategies for Mass-Discharge Estimation". J. Environ. Engineer. Div., ASCE 109(4), 812-829, 1983.

[Bunn 1984] Bunn, D. W., "Applied Decision Analysis". McGraw-Hill, New York, 1984.

[Couillard] Couillard, D. and Lefebvre, Y., "Analysis of Water Quality Indices". J. Environ. Man., 21, 161-179, 1985.

[CRC 1981] CRC, "Handbook of Chemistry and Physics". CRC Press, Boca Raton Fla., 1981.

[CWQG] Canadian Council of Resource and Environment Ministers, "Canadian Water Quality Guidelines". Government Publication, 1987.

[Dinius 1987] Dinius, S.H., "Design of an Index of Water Quality". Water Res. Bul.,AWRA 23(5) 833-843, 1987.

[Forgey 1985] Forgey, C.L., "OPS83 User's Manual and Report". Production Systems Technologies inc, 1985.

[Inhaber 1975] Inhaber, H., "Environmental Indices". John Wiley & Sons, New York, 1976.

[Hart 1986] Hart, A.,"Knowledge Acquisition for Expert Systems". Kogan Page, London, 1986.

[Kidd 1987] Kidd, A.L., "Knowledge Acquisition for Expert Systems". Plenum Press, New York, 1987.

[Loftis & Ward] Loftis, J.C., and Ward. R.C., "Evaluating Stream Standard Violations Using a Water Quality Data Base". Water Resources Bulletin, AWRA, 17(6), 1071-1077, 1981.

[MOE 1987] Effluent Monitoring Priority Pollutants List, MOE publication, 1987.

[MOE 1986] Eddy, J., "BEALE Ratio Estimator for Pollutant Load Calculations". Unpublished MOE Report, 1986.

[MOE 1984] MOE, "Water Management Goals, Policies, Objectives and Implementation Procedures of the Ministry of the Environment". MOE, 1984.

[Mumford 1985] Mumford, E.,"Knowledge Engineering, Expert Systems and Organizational Change". Manchester Business School, 1985.

[Tore 1987] Tore, A., "Logic Programming and Knowledge Engineering", Addison-Wesley, Don Mills, 1987.

[Truett 1975] Truett, J.B. et al., "Development of Water Quality Management Indices". Water Res. Bul., AWRA 11(3), 436-448, 1975.

[Verschueren] Verschueren, K.,"Handbook of Environmental Data on Organic Chemicals". Van Nostrand, Toronto, 1977.

[Ward & Loftis] Ward, R.C., and Loftis, J.C.,"Establishing Statistical Design Criteria for Water Quality Monitoring Systems: Review and Synthesis". Water Resources Bulletin,AWRA, 22(5), 759-767, 1986.